



## Research Paper

## Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation

Gavin M. Bidelman<sup>a, b, c, \*</sup>, Mary Katherine Davis<sup>a</sup>, Megan H. Pridgen<sup>a</sup><sup>a</sup> School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA<sup>b</sup> Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA<sup>c</sup> University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

## ARTICLE INFO

## Article history:

Received 25 January 2018

Received in revised form

18 May 2018

Accepted 23 May 2018

Available online 26 May 2018

## Keywords:

Auditory scene analysis

Distributed source analysis

Event-related potentials (ERPs)

Frequency-following response (FFR)

Speech processing

Functional connectivity

## ABSTRACT

Everyday speech perception is challenged by external acoustic interferences that hinder verbal communication. Here, we directly compared how different levels of the auditory system (brainstem vs. cortex) code speech and how their neural representations are affected by two acoustic stressors: noise and reverberation. We recorded multichannel (64 ch) brainstem frequency-following responses (FFRs) and cortical event-related potentials (ERPs) simultaneously in normal hearing individuals to speech sounds presented in mild and moderate levels of noise and reverb. We matched signal-to-noise and direct-to-reverberant ratios to equate the severity between classes of interference. Electrode recordings were parsed into source waveforms to assess the relative contribution of region-specific brain areas [i.e., brainstem (BS), primary auditory cortex (A1), inferior frontal gyrus (IFG)]. Results showed that reverberation was less detrimental to (and in some cases facilitated) the neural encoding of speech compared to additive noise. Inter-regional correlations revealed associations between BS and A1 responses, suggesting subcortical speech representations influence higher auditory-cortical areas. Functional connectivity analyses further showed that directed signaling toward A1 in both feedforward cortico-collicular (BS→A1) and feedback cortico-cortical (IFG→A1) pathways were strong predictors of degraded speech perception and differentiated “good” vs. “poor” perceivers. Our findings demonstrate a functional interplay within the brain’s speech network that depends on the form and severity of acoustic interference. We infer that in addition to the quality of neural representations within individual brain regions, listeners’ success at the “cocktail party” is modulated based on how information is transferred among subcortical and cortical hubs of the auditory-linguistic network.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

In natural listening environments, noise and reverberation hinder the successful extraction of speech information (for review, see Bidelman, 2017). Although both are acoustic interferences, each has a distinct effect on speech signals (Helfer and Wilber, 1990; Nabelek and Dagenais, 1986). Noise is caused by the addition of external competing sounds that mask target speech. Contrastively, reverberation is the persistence of reflected acoustic energy in the sound field caused by internal room acoustics (Kinsler et al., 2000). Reverberation produces an overlap between direct and indirect

sounds that “smears” a signal’s spectrum. Natural reverberation also acts to low-pass filter speech compared to the overall flattening of modulations produced by noise. Conveniently, the degree of noise and reverberation superimposed on a target signal can be quantified by similar metrics. For noise, the relative contribution of “noise” and “signal” are characterized via the signal-to-noise ratio (SNR). Similarly, the proportions of acoustic energy attributable to signal and reverberant energy are characterized (in dB) by the direct-to-reverberant energy ratio (D/R), or less commonly, “wet-to-dry” ratio (von Békésy, 1938; Zahorik, 2002).

While both acoustic stressors hinder intelligibility, behavioral studies reveal that human listeners show a differential sensitivity when perceiving signals in noise vs. reverberation (Larsen et al., 2008; McShefferty et al., 2015; Zahorik, 2002). That is, even when the relative intensities between signal and interference are matched in severity (i.e.,  $SNR \approx D/R$ ), noise and reverberation

\* Corresponding author. School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN, 38152, USA.

E-mail address: [g.bidelman@memphis.edu](mailto:g.bidelman@memphis.edu) (G.M. Bidelman).

impact speech perception in different manners. While each interference reduces speech understanding by ~15–20%, vowel confusion patterns can differ under these two acoustic backdrops (Nabelek and Dagenais, 1986). Interestingly, reverb also induces less listening effort than noise during speech comprehension, even at similar levels of behavioral performance (Picou et al., 2016). This suggests that while there is a comparable tax on speech intelligibility, noise and reverberation might uniquely impact the underlying neural representations for speech (cf. White-Schwoch and Kraus, 2017). To our knowledge, this possibility has not been formally tested.

It is now well-established that speech-in-noise (SIN) understanding is determined by more than audibility or peripheral hearing status (Middelweerd et al., 1990; Song et al., 2011) (but see Humes and Roberts, 1990). The fact that SIN performance is not reliably predicted from the audiogram (Killion and Niquette, 2000) and varies among even normal-hearing listeners (Song et al., 2011) has led to the notion that central brain mechanisms play a critical role in supporting successful cocktail party listening (e.g., Alain et al., 2014; Bidelman and Howell, 2016; Billings et al., 2009; Billings et al., 2010; Billings et al., 2013; Parbery-Clark et al., 2011; Song et al., 2011). In this regard, electrophysiological studies have been important in elucidating the central factors of speech sound processing.

The brain's neuroelectric response to speech reflects an aggregate of activity generated from both brainstem and cortical structures. By exploiting properties of each response (e.g., spectral bandwidth), one can isolate their contributions within the scalp EEG and examine sound encoding within various structures of the auditory hierarchy (Bidelman et al., 2013). The cortical ERPs are composed of several "waves" (e.g., P1-N1-P2), reflecting activation of auditory thalamus, cortex, and associative areas (Picton et al., 1999). ERPs are sensitive to the acoustic features of speech (Agung et al., 2006; Chang et al., 2010; Kraus and Cheour, 2000; Sharma and Dorman, 1999) and correlate with listeners' perceptual skills (Bidelman et al., 2014b; Ross and Tremblay, 2009; Tremblay et al., 2001). The *subcortical* component, or frequency-following response (FFR), is a sustained potential emitted dominantly from the upper brainstem that closely mirrors acoustic stimuli with high fidelity (Bidelman, 2018; Krishnan, 2007; Skoe and Kraus, 2010). FFR activity similarly correlates with listeners' SIN perception (Anderson et al., 2010; Parbery-Clark et al., 2009a; Song et al., 2011). Yet, few studies have examined FFRs to reverberant speech (Bidelman and Krishnan, 2010; Fujihira and Shiraishi, 2015), and we unaware of any directly contrasting the effects of noise and reverb on speech FFRs. Moreover, while a number of studies have investigated the independent contributions of brainstem (e.g., Bidelman, 2016; Bidelman and Krishnan, 2010; Billings et al., 2013; Krishnan et al., 2010; Parbery-Clark et al., 2009a; Song et al., 2011) and cortical neurophysiology (e.g., Alain et al., 2014; Bidelman and Howell, 2016; Billings et al., 2010; Billings et al., 2013; Shtyrov et al., 1998) to degraded speech processing, examining these functional levels simultaneously (within individual listeners) would offer a more comprehensive, systems-level characterization of the biological mechanisms underlying cocktail party listening in different acoustic scenarios and possible interplay between stages of the neuroaxis.

To this end, our recent studies have championed the use of concurrent FFR-ERPs recordings to examine hierarchical auditory processing (e.g., Bidelman, 2015; Bidelman and Alain, 2015b; Bidelman et al., 2013; Bidelman et al., 2014b) (see also Bellier et al., 2015; Slugocki et al., 2017). Dual FFR-ERP paradigms have provided important insight into how lower vs. higher tiers of the neuroaxis code complex sounds and interact during early perception when object-based representations of speech are still in their nascent

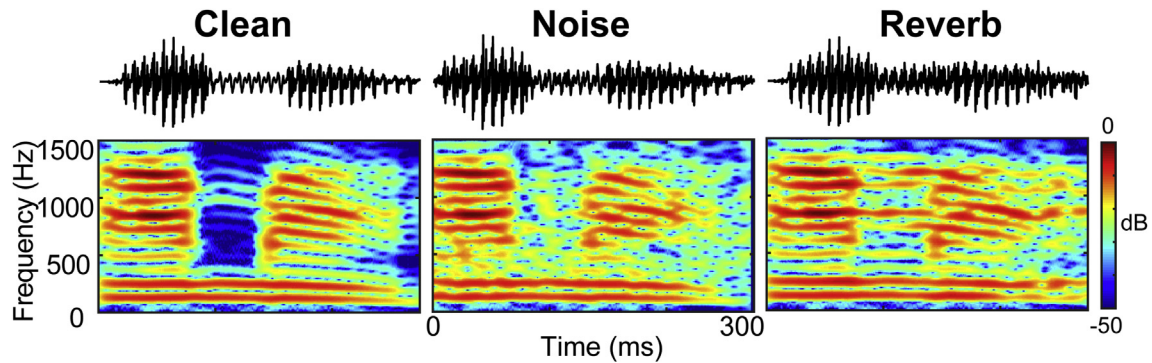
stages (e.g., Bidelman et al., 2013). Translational applications have further shown how brainstem vs. cortical functions are uniquely strengthened by plasticity (Bidelman and Alain, 2015a; Bidelman et al., 2014a; Musacchia et al., 2008), are differentially compromised by hearing loss (Bidelman et al., 2014b), and are altered in neurocognitive disorders (Bidelman et al., 2017). This approach emphasizes a growing body of work that advocates speech processing as an integrative and dynamic network (Kraus and White-Schwoch, 2015; Obleser et al., 2007; Scott et al., 2009) which includes possible interactions and/or signal transformations between brainstem-cortex (Gao and Suga, 1998; Suga et al., 2002) and auditory-linguistic brain areas downstream (Du et al., 2014). Given that early brainstem-cortical and later auditory-linguistic interplay can presumably vary on an individual basis, we hypothesized these individual differences might modulate cocktail party perception. Here, we exploited dual FFR-ERPs to further investigate the neural encoding of impoverished (noisy and reverberant) speech and define the functional connectivity between subcortical and cortical hubs of the auditory system.

The aims of the present study were thus twofold. First, we directly compared how subcortical and cortical levels of the auditory system encode different forms of degraded speech information. While previous reports have investigated relations between brainstem and cortical auditory processing (e.g., Bidelman et al., 2013; Bidelman et al., 2017; Krishnan et al., 2012; Musacchia et al., 2008; Slugocki et al., 2017), studies have focused exclusively on scalp (electrode-level) recordings and therefore, can only *infer* contributions of sub- and neo-cortex from volume conducted mixtures of EEG signals. Here, source analysis of brainstem FFRs and cortical ERPs allowed us to parse region-specific activity with higher granularity and more definitively reveal how neural processing *within* each tier of the neuroaxis coordinate during speech processing. Functional connectivity analysis evaluated the directed, causal signaling between brainstem and cortical regions and how inter-regional neural communication might predict listeners' speech perception skills. We also measured source FFR/ERPs elicited by noisy and reverberant speech. This allowed us to directly assess how neural speech representations are affected by different acoustic stressors common to the auditory scene. To anticipate, our data reveal that degraded speech perception is governed by intra- and inter-regional brainstem-cortical activity including cortico-collicular (brainstem-cortical) and cortico-cortical (fronto-temporal) signaling.

## 2. Methods

### 2.1. Participants

Eleven adults (age:  $24.7 \pm 2.7$  years) participated in the experiment. All had obtained a similar level of formal education (at least an undergraduate degree), and were monolingual speakers of American English. Musical training is known to enhance auditory evoked responses (e.g., Bidelman et al., 2011; Musacchia et al., 2008; Zendel and Alain, 2009) and improve degraded speech-listening skills (Bidelman and Krishnan, 2010; Parbery-Clark et al., 2009a, 2009b). Hence, all participants were required to have <3 years formal musical training ( $1.3 \pm 1.8$  years) and none within the past 5 years. Audiometric screening confirmed normal hearing (i.e., thresholds < 25 dB HL) at octave frequencies (250–8000 Hz). All participants reported no history of neuropsychiatric disorders. Participants gave written informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board and were compensated monetarily for their time.



**Fig. 1. Speech stimuli.** Example time waveforms and spectrograms for the speech stimulus /ama/ presented in clean (no noise), noise-degraded (+5 dB SNR), and reverberant (+5 dB D/R) conditions.

## 2.2. Stimuli

Brainstem and cortical auditory ERPs were elicited by a 300 ms /vCv/ speech token /ama/ (Bidelman and Howell, 2016) (Fig. 1). The pitch prosody fell gradually over the token's duration ( $F_0 = 120 \text{ Hz} - 88 \text{ Hz}$ ). Vowel formant frequencies ( $F_1 - F_3$ ) were 830, 1200, and 2760 Hz, respectively.

**Noise-degraded speech.** In addition to this “clean” stimulus (SNR =  $\infty$  dB), noise-degraded stimuli were created by adding four-talker babble noise (Killion et al., 2004) to the clean token at SNRs of +10 and +5 dB. SNR was manipulated by changing the level of the masker rather than the signal to ensure SNR was inversely correlated with overall sound level (Binder et al., 2004). To mimic real-world acoustic scenarios, babble was presented continuously so that it was not time-locked to the stimulus presentation (e.g., Alain et al., 2012).

**Reverberant speech.** Reverberant speech was simulated via convolution reverb. We used the XFX1 (v1) reverb plugin in Sound Forge 9 (Sony Inc.) with the “Rich Hall” preset to simulate acoustics of a realistic room ( $RT_{60}$ : 500 ms; early reflection style: 16 ms; pre-delay: 30 ms; filter attenuation: 20–12,000 Hz; early reflection attenuation:  $-20$  dB). We then exported the “wet” (reverberant) portion of the signal (i.e., 0% mix of the “dry” speech) to extract only the “interference” component of the reverberant signal. Effectively, this resulted in a simulation as if the speech was produced in a reverberant environment but with only indirect (reflected) energy containing no direct sound. We then mixed the original (clean) speech token with this reverberation to create reverb-degraded speech at mild and medium direct-to-reverberant ratios (D/R) of +10 and +5 dB, respectively. Importantly, this approach allowed us to precisely control and equate the overall level of severity (i.e., SNR and D/R) between noise- and reverb-degraded stimuli. This avoided potential confounds of having different degrees of acoustic “noisiness” (i.e., signal figure-ground quality) between noise and reverb conditions and allowed us to focus on differences in sound quality produced by each acoustic stressor.<sup>1</sup> Acoustic waveform and spectrograms are illustrated for the clean, +5 dB SNR noise-degraded, and +5 dB D/R reverberant speech in Fig. 1.

<sup>1</sup> While figure-ground quality was equated between noisy and reverberant speech (i.e., SNR = D/R), we did not attempt to control for inherent spectrotemporal differences between interference classes. For example, unlike additive noise, reverb distorts the target itself as reflected portions of the speech signal overlap with itself and lead to a blurring or “temporal smearing” of the waveform's fine structure (Nabelek et al., 1989; Wang and Brown, 2006). Thus, while we control the overall level of interference, we retain intrinsic differences that are characteristic of noise vs. reverberation (e.g., timbral flux, dynamic temporal effects, lowpass filtering in reverb vs. overall flattening of modulation in noise).

## 2.3. Procedure

Listeners heard 2000 trials of each condition (passive listening) presented with fixed, rarefaction polarity using a jittered inter-stimulus interval (400–700 ms; 20-ms steps, uniform distribution). Stimulus presentation was controlled by MATLAB® 2013b (The MathWorks) routed to a TDT RP2 interface (Tucker-Davis Technologies). Tokens were delivered binaurally at 81 dB SPL through ER-30 insert earphones (Etymotic Research). ER-30 earphones feature an extended acoustic tubing (20 ft) that allowed us to place their transducers outside the testing booth. Spatially separating the transducer from participants avoided the possibility of electromagnetic stimulus artifact from contaminating neural responses (Aiken and Picton, 2008; Campbell et al., 2012). The low-pass frequency response of the headphone apparatus was corrected with a dual channel 15 band graphical equalizer (dbx EQ Model 215s; Harman) to achieve a flat frequency response through 4 kHz. Stimulus level was calibrated using a Larson–Davis SPL meter (Model LxT) measured in a 2-cc coupler (IEC 60126). The entire experimental procedure, including electrophysiological and behavioral testing (described below) took ~2 h.

## 2.4. Behavioral speech-in-noise task

Before EEG testing, we measured listeners' speech reception thresholds in noise using the QuickSIN test (Killion et al., 2004). Participants were presented two lists of six sentences with five key words per sentence embedded in four-talker babble noise. Sentences were presented at 70 dB SPL via insert earphones (bilaterally) using pre-recorded signal-to-noise ratios (SNRs) that decreased in 5 dB steps from 25 dB (very easy) to 0 dB (very difficult). Listeners scored one point for each key word correctly repeated. “SNR loss” (computed in dB) was determined as the SNR required to correctly identify 50% of the key words (Killion et al., 2004). We averaged SNR loss from two lists for each listener. As there is no reverb analog to the QuickSIN, this test functioned as a (normative) behavioral assay of degraded speech listening skills.

## 2.5. Electrophysiological recordings

**EEG acquisition and preprocessing.** Electrophysiological recording procedures were identical to our previous report (Bidelman and Howell, 2016). Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 scalp locations (Oostenveld and Praamstra, 2001). EEGs were digitized at a sampling rate of 5000 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC–2500 Hz. 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. Data were re-referenced off-line to a common average reference for subsequent analyses. Contact impedances were maintained  $\leq 5$  k $\Omega$ .

Subsequent pre-processing was performed in Curry 7 (Compumedics Neuroscan). Ocular artifacts (saccades and blinks) were first corrected in the continuous EEG using a principal component analysis (PCA) (Picton et al., 2000). Cleaned EEGs were then epoched (–200–550 ms), baseline-corrected to the pre-stimulus period, and averaged in the time domain to obtain compound evoked responses (containing both brainstem and cortical activity) for each stimulus condition per participant.

**Source waveform derivations.** Scalp potentials (sensor-level recordings) were transformed to source space using discrete inverse models. Fitting was carried out in BESA<sup>®</sup> Research v6.1 (BESA, GmbH). We used a four-shell volume conductor head model (Berg and Scherg, 1994; Sarvas, 1987) with BESA default settings, i.e., relative conductivities (1/ $\Omega$ m) of 0.33, 0.33, 0.0042, and 1 for the head, scalp, skull, and cerebrospinal fluid, respectively, and sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness) (Herdman et al., 2002; Picton et al., 1999). From grand averaged responses, we seeded dipoles in prominent hubs of the auditory-speech network, including brainstem (BS), bilateral primary auditory cortex (A1), and bilateral inferior frontal gyrus (IFG). Although a simplistic speech circuit, these three regions of interest (ROIs) have been shown to predict SIN processing in several neuroimaging studies (Adank et al., 2012; Bidelman and Krishnan, 2010; Bidelman and Dexter, 2015; Bidelman and Howell, 2016; Binder et al., 2004; Coffey et al., 2017; Du et al., 2014; Scott and McGettigan, 2013; Song et al., 2011) and allowed specific hypothesis testing of the relation between auditory brainstem-cortical connectivity and degraded speech perception. Sparse source models have also been shown to better resolve subcortical sources in M/EEG than denser dipole configurations (Krishnaswamy et al., 2017), further justifying a restricted dipole count. While EEG inverse solutions are non-unique, we have confirmed the biological plausibility of this brainstem-cortical configuration (clean spatial separation, spatiotemporal dynamics) using identical modeling assumptions in our recent source imaging and computational studies (Bidelman, 2018). Overall, the model provided an excellent fit to the grand averaged data across stimuli and listeners (goodness of fit:  $82 \pm 4.5\%$ ). This dipole configuration was then used to guide source derivations at the single-subject level.

To extract each individual's source waveforms within each ROI, we transformed listener's scalp recordings into source-level responses using a virtual source montage (Scherg et al., 2002). This digital re-montaging applies a spatial filter to all electrodes (defined by the foci of our dipole configuration). Relative weights are then optimized to image activity within each brain ROI while suppressing overlapping activity stemming from other active brain regions (for details, see Scherg and Ebersole, 1994; Scherg et al., 2002). This allowed us to reduce each listener's electrode recordings (64-channels) to 5 source channels, each of which estimated activity generated within a single ROI (i.e., left/right IFG, left/right A1, BS). For each participant, the model was held fixed and was used as a spatial filter to derive their source waveforms (Zendel and Alain, 2014), reflecting the neuronal current (in units nAm) as seen within each anatomical ROI. Source waveforms were then bandpass filtered into high- (80–1500 Hz) and low- (0.5–20 Hz) frequency bands to isolate the periodic brainstem FFR vs. slower cortical ERP waves from each listeners' compound evoked potential (Bidelman et al., 2013; Musacchia et al., 2008). Comparing FFR and ERP source waveforms allowed us to assess the relative

contribution of brainstem and cortical generators to degraded speech processing. Identical model assumptions for the inverse solution (e.g., volume conductor, conductivities) were used for brainstem and cortical ROIs waveforms; only the anatomical location of the dipole differed. Importantly, EEG provides more consistent SNR for both deep (brainstem) and superficial sources (cortex) and between focal and distributed models compared to other neuroimaging modalities (e.g., MEG; Goldenholz et al., 2009).

**FFR/ERP source waveform quantification.** FFRs and ERP source waveforms were quantified based on their amplitude (FFR: RMS amplitude; ERP: N1 amplitude). Previous studies have shown correspondence between these measures and degraded speech perception (Bidelman and Howell, 2016; Billings et al., 2013; Parbery-Clark et al., 2011). The cortical N1 was taken as the negative-going deflection between 100 and 150 ms based on the expected noise-related delay in the ERPs (Bidelman and Howell, 2016; Billings et al., 2009) after factoring in the acoustic delay of the headphones based on the speed of sound in air [17.8 ms = (20 ft)/(1125 ft/s)].

## 2.6. Distributed source imaging (CLARA)

We used Classical Low Resolution Electromagnetic Tomography Analysis Recursively Applied (CLARA) [BESA (v6.1)] to estimate the neuronal current density underlying the sensor data recorded at the scalp. Distributed analyses were applied only to cortical activity (i.e., 0.5–20 Hz filtered data). The CLARA technique models the inverse solution as a large collection of elementary dipoles distributed over nodes on a mesh of the cortical volume. The aggregate strength of current density in each voxel can then be projected spatiotemporally onto the neuroanatomy, akin to a functional map in fMRI. CLARA estimates the total variance of the scalp-recorded data and applies a smoothness constraint to ensure current changes little between adjacent brain regions (Michel et al., 2004; Picton et al., 1999). CLARA renders highly focal source images by iteratively reducing the source space during repeated estimations. On each step ( $\times 10$ ), a spatially smoothed LORETA solution (Pascual-Marqui et al., 2002) is recomputed and voxels below a 1% max amplitude threshold are removed. This provided a spatial weighting term for each voxel of the LORETA image on the subsequent step. Ten iterations were used with a voxel size of 5 mm in Talairach space and regularization (parameter accounting for noise) set at 0.01% singular value decomposition. The spatial resolution of CLARA is estimated at 5–10 mm (Iordanov et al., 2014, 2016), which is smaller (better) than the physical distances between PAC and IFG, and between BS and PAC (~35–40 mm) (Mazziotta et al., 1995). We have previously used this approach to image electrophysiological speech processing (e.g., Alain et al., 2017; Bidelman, 2018). Source activations were visualized on BESA's adult brain template (Richards et al., 2016).

We used cluster-based permutation tests (Maris and Oostenveld, 2007) implemented in BESA Statistics (2.0) to contrast the distributed source images and identify anatomical locations in the brain volume (and over time) that distinguished noise-vs. reverb-degraded speech. For each voxel and time point, a paired samples *t*-test was conducted contrasting the two mild forms of acoustic interference (i.e., +10 dB SNR noise vs. +10 D/R reverb). This allowed us to determine where and when neural activity showed significant differences ( $p < 0.05$ ) between interferences. Cluster values were derived based on the sum of all *t*-values of data points within a given cluster. Significant differences were determined by generating and comparing surrogate clusters from  $n = 1000$  resamples of the data permuting between stimulus conditions (e.g., Oostenveld et al., 2011). This identified contiguous clusters of voxels over time (i.e., statistical contrast

maps) where stimulus conditions were not interchangeable (i.e., noisy speech  $\neq$  reverberant speech;  $p < 0.05$ ). Importantly, BESA corrects for multiple comparisons across the aggregate of all voxels and time points by controlling the familywise error rate through this clustering process. The same procedure was then repeated to contrast the two medium forms of interference (i.e., +5 dB SNR noise vs. +5 D/R reverb). In the present study, distributed source imaging was used to identify qualitatively, the most relevant cortical regions that distinguish noise and reverberant speech. Source montages, based on discrete dipole modeling, were then used to quantitatively investigate the time course of speech-evoked responses within each cortical ROI and assess their relation to brainstem activity.

### 2.7. Functional connectivity

We measured causal (directed) information flow between nodes of the brainstem-cortical speech network using phase transfer entropy (PTE) (Lobier et al., 2014). TE is a non-parametric, information theoretic measure of directed signal interaction. It is ideal for measuring functional connectivity between regions because it can detect nonlinear associations between signals and is robust against the volume conducted crosstalk in EEG (Hillebrand et al., 2016; Vicente et al., 2011). TE was estimated using the time series of the instantaneous phases of the signals, yielding the so-called phase transfer entropy (PTE) (Hillebrand et al., 2016; Lobier et al., 2014). PTE was computed between ROIs according to Eq. (1):

$$PTE_{X \rightarrow Y} = \sum p(y_{t+\tau}, y_t^m, x_t^n) \log_2 \left[ \frac{p(y_{t+\tau} | y_t^m, x_t^n)}{p(y_{t+\tau}, y_t^m)} \right] \quad 1$$

where  $X$  and  $Y$  are the ROI signals and the  $\log(\cdot)$  term is the conditional probabilities between signals at time  $t+\tau$  for sample  $m$  and  $n$ . The probabilities were obtained by building histograms of occurrences of pairs of phase estimates in the epoch (Lobier et al., 2014). Following Hillebrand et al. (2016), the number of histogram bins was set to  $e^{0.626+0.4 \ln(N_s - \tau - 1)}$  (Otnes and Enochson, 1972). The prediction delay  $\tau$  was set as  $(N_s \times N_{ch}) / N_{\pm}$ , where  $N_s$  and  $N_{ch}$  are the number of samples and ROI sources, respectively, and  $N_{\pm}$  the number of times the phase changes sign across time and ROI signals (Hillebrand et al., 2016). PTE cannot be negative and has no upper bound. Higher values indicate stronger connectivity, whereas  $PTE_{X \rightarrow Y} = 0$  implies no directed signaling.

Intuitively, PTE can be understood as the reduction in information (in units bits) necessary to encode the present of ROI<sub>Y</sub> if the past of ROI<sub>X</sub> is used in addition to the past of ROI<sub>Y</sub>. In this sense, it is similar to the definition of Granger Causality (Barnett et al., 2009), which states that ROI<sub>X</sub> has a causal influence on the target ROI<sub>Y</sub> if knowing the past of both signals improves the prediction of the target's future compared to knowing only its past. Yet, PTE has several important advantages over other connectivity metrics: (i) PTE is more robust to realistic amounts of noise and linear mixing in the EEG that can produce false-positive connections; (ii) PTE relaxes assumptions about data normality and is therefore model-free; (iii) PTE is asymmetric so it can be computed bi-directionally between pairs of sources ( $X \rightarrow Y$  vs.  $Y \rightarrow X$ ) to infer causal, directional flow of information between interacting brain regions. Computing PTE in both directions between pairwise sources (e.g., BS  $\leftrightarrow$  A1 and A1  $\leftrightarrow$  IFG) allowed us to quantify the relative weighting of information flowing between subcortical and cortical ROIs in both feedforward and feedback directions.

## 3. Results

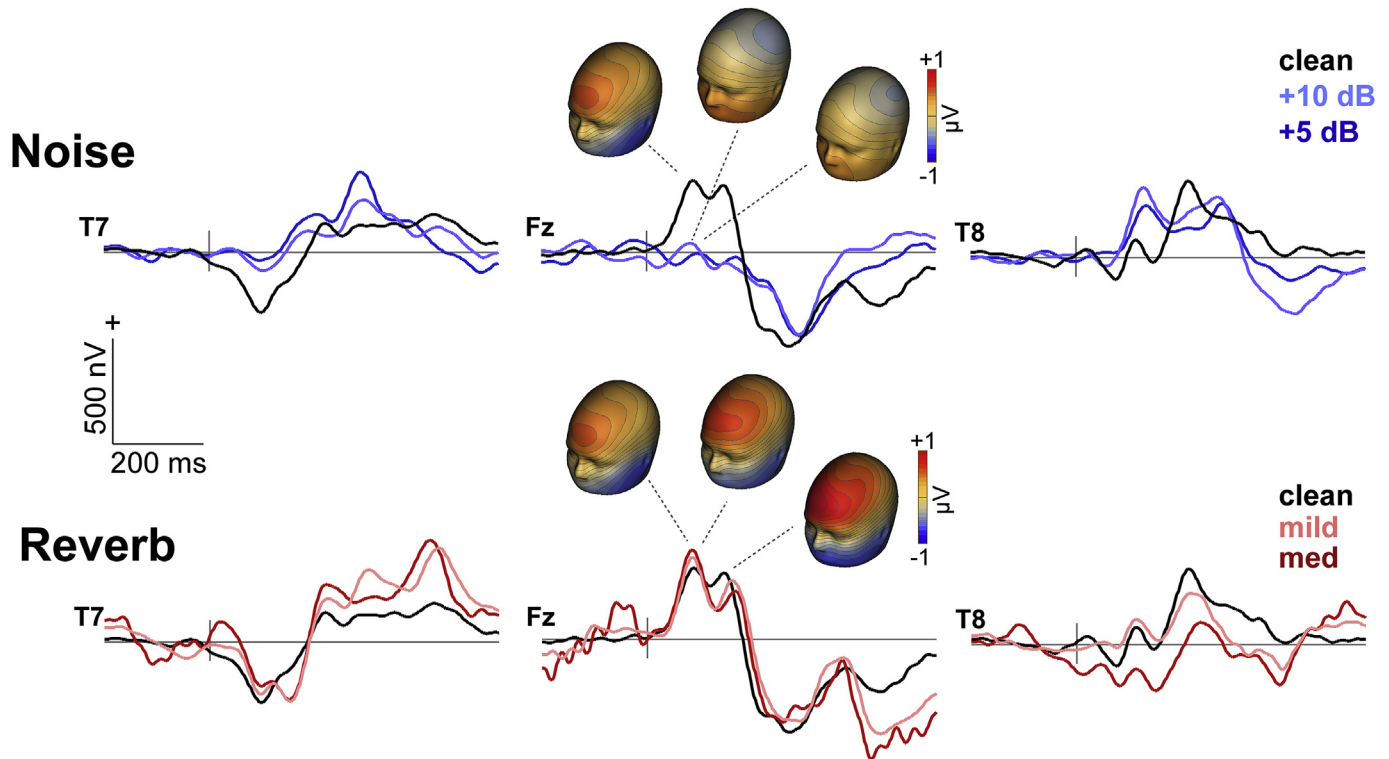
### 3.1. Sensor-level data

Scalp-recorded cortical ERPs and brainstem FFRs (electrode data) to noise and reverberant speech are shown in Figs. 2–3. Visual inspection of scalp maps suggested weaker cortical responses with decreasing SNR for noise-degraded speech (Fig. 2). Weaker N1-P2 is presumably due to the decrease in inter-trial response coherence (i.e., increased jitter) caused by noise (Koerner and Zhang, 2015). This amplitude reduction was not observed for reverberant speech where ERPs showed little apparent change with increasing severity of reverb. Similarly, scalp FFRs (Fig. 3), reflecting phase-locked brainstem activity, showed a reduction in amplitude with increasing levels of noise but an apparent enhancement for speech in reverberation. This was most apparent when directly comparing mild (+10 dB SNR noise vs. mild reverb) and moderate interferences (+5 dB SNR noise vs. med. reverb). Despite matching noise SNR and reverb D/R, reverberant speech produced larger responses, particularly at the voice fundamental frequency (F0) (Fig. 3B). The volume conducted nature of sensor recordings did not allow for the separation of underlying sources. Hence, subsequent analyses were conducted in source space.

### 3.2. Source waveform data

CLARA distributed source maps contrasting cortical activity (0.5–20 Hz) to noise- and reverb-degraded speech are shown in Fig. 4. Contrasts between the +10 dB SNR noise and mild reverb conditions revealed responses to these mild forms of interference were differentiated in bilateral, but especially right, auditory cortex (Fig. 4A). The maximal effect (i.e., reverb > noise) occurred 80 ms after stimulus onset. Similarly, contrasts between the +5 dB SNR noise and medium reverb conditions revealed that responses to moderate interferences were differentiated ~30 ms later (108 ms), primarily within left IFG (Fig. 4B). These data suggest a distributed cortical network for speech consisting of auditory (A1) and linguistic (IFG) brain regions that is differentially engaged, intra- and inter-hemispherically, depending on the severity of acoustic interference. While these distributed source analyses provided a whole-brain view of degraded speech processing, they do not allow specific hypothesis testing within the network. Consequently, we extracted source waveforms within cortical A1 and IFG as well as brainstem ROIs to investigate how noise and reverb modulate the subcortical-cortical encoding of speech. Following our previous report (Bidelman and Howell, 2016), left and right hemisphere responses were collapsed for subsequent analyses. This reduced the dimensionality of the data and provided increased SNR for source-level analysis. Pooling hemispheres is also motivated by the fact that even monaural stimulus presentation produces bilateral auditory cortical activity (Schonwiesner et al., 2007).

Brainstem (BS) and cortical (A1, IFG) source waveforms are shown in Fig. 5A and B, respectively. BS sources illustrate the periodic frequency-following response (FFR) of the auditory midbrain (Bidelman, 2018). Response spectra (lowest panels) illustrate phase-locked activity to the voice pitch (F0  $\approx$  120 Hz) and harmonic structure of speech. A1 and IFG waveforms track the time course of speech encoding in primary auditory and linguistic brain regions and modulations with noise and reverberation. We used a mixed model, 3-way ANOVA (ROI  $\times$  interference type  $\times$  interference level; subject = random factor) to analyze response amplitudes. Initial diagnostics revealed that amplitudes were leptokurtic (i.e., high peakedness) in their distribution. Hence, amplitudes measures were  $\text{asinh}(\cdot)$  transformed to improve normality and homogeneity of variance assumptions necessary for parametric statistics. The



**Fig. 2. Cortical ERPs (electrode data).** Grand average responses at frontal and temporal channels (T7/8, Fz) to speech as a function of noise SNR (top traces) and reverb D/R (bottom traces). Responses appear as a series of obligatory waves within the first 200 ms of speech characteristic of the auditory P1-N1-P2 signature. Topographies illustrate the distribution of activity across the scalp. Blue map colors refer to negative voltage, red colors indicate positive voltage. Reverberation is less detrimental to (and in some cases facilitates) the neural encoding of speech compared to additive noise. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

ANOVA revealed significant main effects of interference type [ $F_{1,140} = 7.01, p = 0.009$ ] and ROI [ $F_{2,140} = 8.45, p = 0.0003$ ]. Planned contrasts revealed linear effects of interference level on BS [ $t_{40} = 3.34, p = 0.0018$ ] and IFG [ $t_{40} = 2.62, p = 0.0124$ ] amplitudes. That is, relative to clean speech, FFR amplitudes steadily increased with additional noise but were also larger for reverberant compared to noisy speech (i.e., Fig. 5C and E; +10 dB < +5 dB < mild < medium). This noise/reverb-related modulation was not observed in A1 [linear effect:  $t_{40} = 0.64, p = 0.53$ ] (Fig. 5D). Within IFG, we also found a significant effect of interference type [ $F_{1,40} = 7.91, p = 0.0076$ ]. That is, despite matched level of severity, speech in mild reverb elicited larger amplitudes than speech in mild noise ( $p = 0.0129$ ). The facilitation effect of reverb on cortical IFG responses parallels the reverb facilitation observed in the sensor-level data (Figs. 2–3).

### 3.3. Relations between brainstem and cortical speech responses

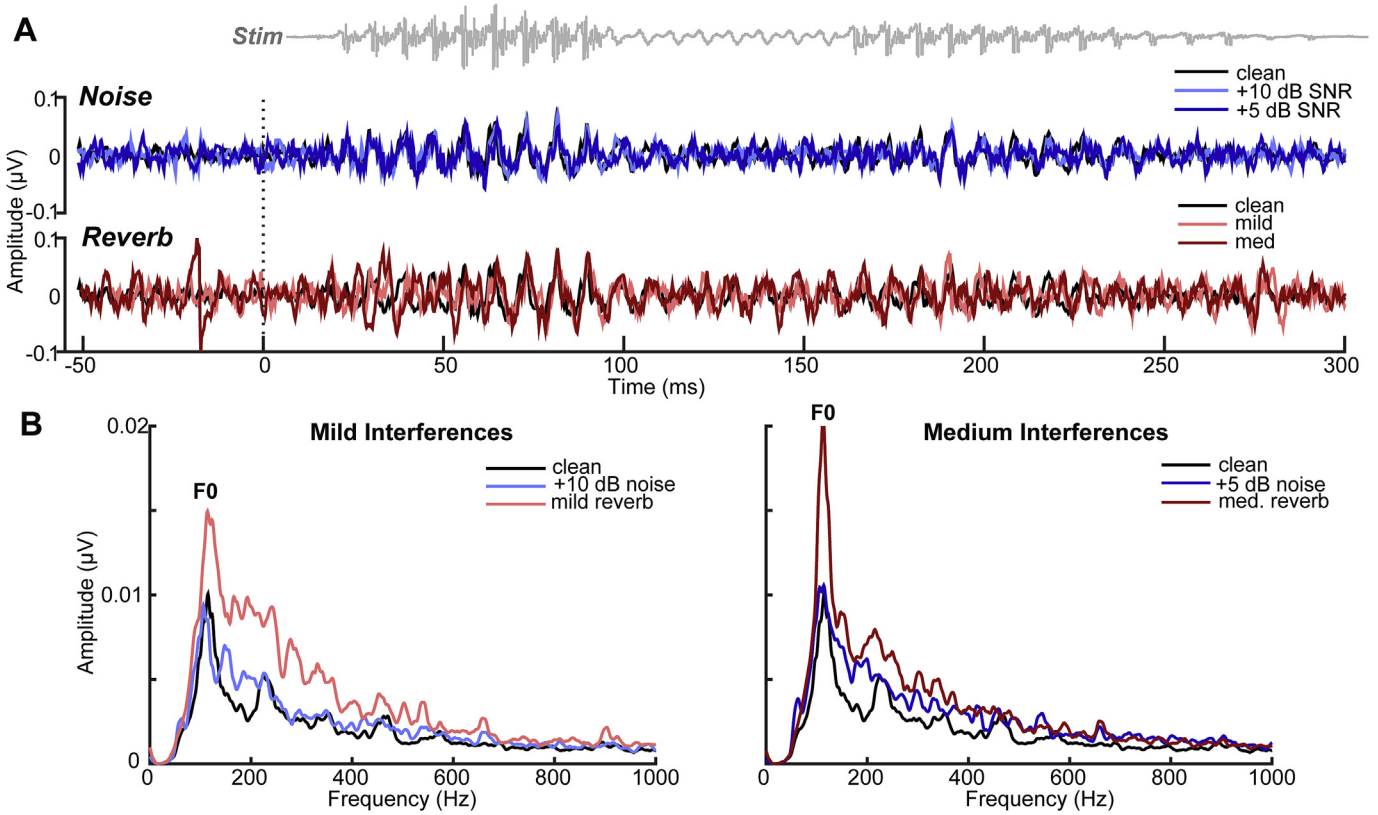
Pairwise Pearson's correlations examined associations between brainstem and cortical responses (Fig. 6). We found that BS-FFRs were positively correlated with A1-ERP amplitudes ( $r = 0.74, p = 0.0089$ ) such that stronger brainstem speech encoding was associated with larger auditory cortical responses (Fig. 6A). This suggests that neural activity along the ascending auditory pathway (BS → A1) might drive responses downstream. No associations were observed between BS and IFG amplitudes (Fig. 6B;  $r = 0.37, p = 0.27$ ) nor between A1 and IFG (Fig. 6C;  $r = 0.20, p = 0.56$ ). The lack of correlation between A1 → IFG (and BS → IFG) implies that speech representations might become more independent/abstract at higher levels of the neural hierarchy.

### 3.4. Brain-behavior relations

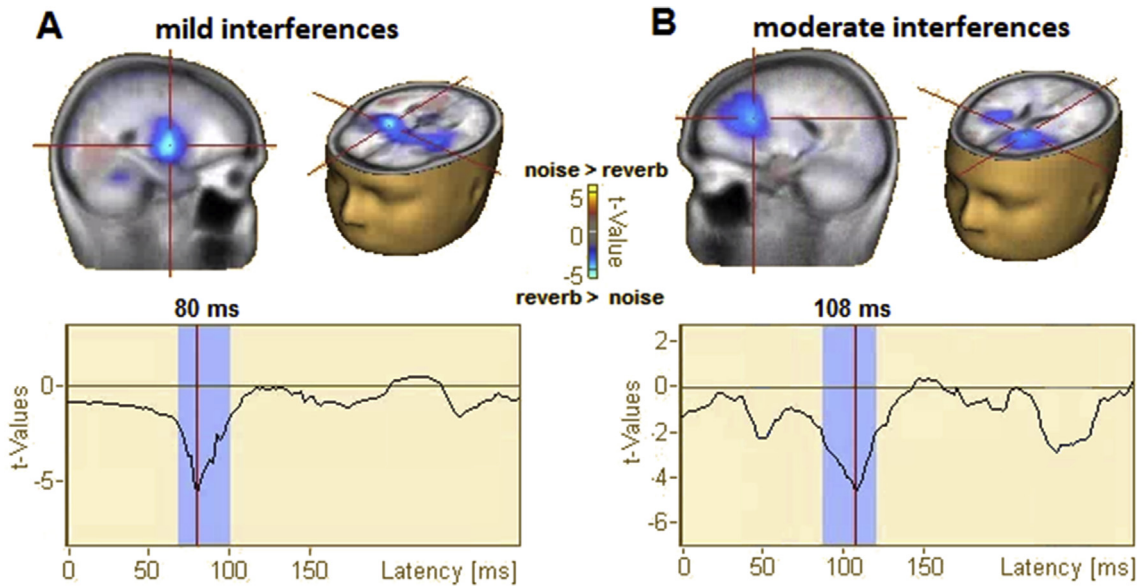
We used a generalized linear mixed effects model (GLME) to evaluate relationships between neural responses to acoustically-degraded speech and behavioral QuickSIN performance (Fig. 7A). We considered the combination of all neural amplitude measures (BS<sub>FFR</sub>, A1<sub>ERP</sub>, IFG<sub>ERP</sub>; pooled across noise/reverb conditions) as well as their interactions as predictors of listeners' QuickSIN scores. Participants were modeled as a random term [i.e., QuickSIN ~ FFR\*A1\*IFG + (1|subject)]. Dependent variables were mean centered to reduce multicollinearity (Afshartous and Preston, 2011). We found that the overall model was significant [ $F_{1,3} = 11.56, p = 0.037$ ; null hypothesis coefficients = 0] accounting for 81% of the variance in QuickSIN scores (adjusted- $R^2 = 0.81$ ). Scrutiny of individual model terms revealed significant neural predictors in BS ( $p = 0.0241$ ) and IFG ( $p = 0.011$ ) amplitudes. In addition to these focal responses, we observed significant A1\*IFG ( $p = 0.018$ ) and BS\*IFG ( $p = 0.011$ ) interactions. No other terms reached significance. Collectively, these results suggest that in addition to intra-regional activity within brainstem and cortical areas (BS, IFG), cortico-collicular (BS-A1) and cortico-cortical (A1-IFG) interactions are strong predictors of degraded speech perception abilities.

### 3.5. Functional connectivity

Phase transfer entropy, quantifying the feedforward (afferent) and feedback (efferent) functional connectivity between sources, is shown in Fig. 7C. To investigate the behavioral relevance of PTE connectivity measures, we divided our cohort in two groups based on a median split of listeners' QuickSIN scores. This resulted in two



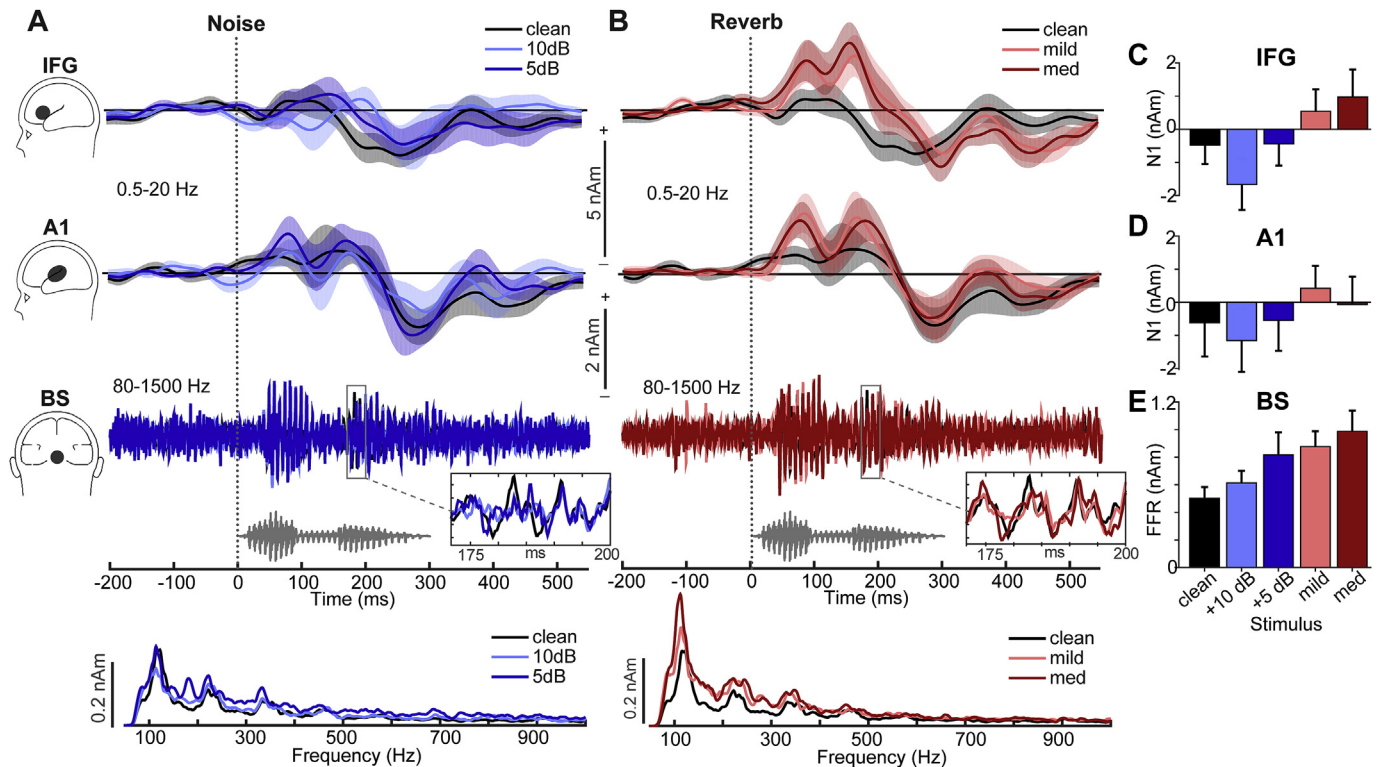
**Fig. 3. Brainstem FFRs (electrode data).** (A) Grand average time waveforms and (B) response spectra to noise-degraded and reverberant speech. F0, voice fundamental frequency. Spectra illustrate enhancements in speech encoding (particularly at F0) with reverberation compared to speech presented in equivalent amounts of interference. Otherwise as in Fig. 2.



**Fig. 4. Statistical contrasts (*t*-maps) of source responses to noise and reverberant speech.** Data are filtered 0.5–20 Hz (i.e., cortical activity). Functional data projected on the brain volume are displayed on the BESA brain template (Richards et al., 2016). (A) Contrast of mild forms of acoustic interference (+10 dB SNR noise vs. mild reverb). (B) Contrast of moderate forms of acoustic interference (+5 dB SNR noise vs. medium reverb). Maps are masked ( $p < 0.05$ ) and corrected for multiple comparisons (see text). Bottom traces show the running *t*-value (noise > reverb) of the cluster marked in the brain volume by crosshairs. Cool colors denote locations where reverberant > noise. Degraded speech is processed in bilateral A1 and IFG. For mild acoustic interferences (A), noisy and reverberant speech are differentiable mainly in right A1. This contrasts speech in moderate interferences (B), which becomes distinguishable ~30 ms later in left IFG. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

subgroups, “good” ( $n = 6$ ) vs. “poor” ( $n = 5$ ) SIN performers who differed in their SIN scores (Fig. 7B). Subgroups were otherwise

matched in age ( $t_9 = 1.61$ ,  $p = 0.14$ ), musical training ( $t_9 = 0.25$ ,  $p = 0.81$ ), and gender (Fisher’s exact test:  $p = 1$ ). Despite all having



**Fig. 5.** Brainstem FFR and cortical ERP source responses to speech in noise and reverberation. Source waveforms (collapsed across hemispheres) extracted from cortical (IFG, A1; 0.5–20 Hz) and brainstem (BS; 80–1500 Hz) dipole sources. (A) Responses in noise. (B) Responses in reverb. Bottom traces below BS-FFR waveforms show the spectra of phase-locked activity from the BS source. Gray trace, stimulus waveform. (C–E) Changes in IFG, A1, and BS source amplitudes across interferences. Shading and errorbars =  $\pm 1$  s.e.m.

normal hearing, good listeners obtained better ( $\sim 1$  dB lower) SIN thresholds [ $t_9 = -2.47$ ,  $p = 0.035$ ], confirming individual differences in SIN perception (Fig. 7B) (e.g., Song et al., 2011). We then analyzed each groups' PTE values to assess whether or not directed, inter-regional neural connectivity (i.e., BS  $\leftrightarrow$  A1 and A1  $\leftrightarrow$  IFG) could distinguish “good” vs. “poor” SIN performers (Fig. 7C). An ANOVA conducted on PTE values revealed a significant group  $\times$  region interaction in the strength of functional connectivity [ $F_{3,27} = 4.79$ ,  $p = 0.0084$ ]. Identical results were also obtained with a Friedman non-parametric ANOVA (group  $\times$  region:  $F_{3,27} = 4.79$ ,  $p = 0.0084$ ). Tukey-Kramer contrasts revealed that this interaction was driven by stronger connectivity in “good” compared to “poorer” perceivers between IFG  $\rightarrow$  A1 ( $p = 0.0041$ ) and marginally stronger connectivity between BS  $\rightarrow$  A1 ( $p = 0.07$ ) (Fig. 7C). No other regional pairs showed group differences in connectivity. These results confirm both bottom up (BS  $\rightarrow$  A1) and top down (IFG  $\rightarrow$  A1) communication directed toward A1 is important for degraded speech perception.

#### 4. Discussion

By measuring source-level FFR and ERPs to noise- and reverberated speech, the present study helps elucidate how central neural representations for communicative signals are affected by different acoustic stressors common to the auditory scene. Collectively, our results relate to three primary findings: (i) reverberation is less detrimental to (and in some cases facilitated) the neural encoding of speech compared to additive noise matched in acoustic severity (i.e., signal figure-ground quality); (ii) listeners' degraded speech perception skills are predicted by intra- and inter-regional brainstem-cortical activity including cortico-colicular (brainstem-cortical) and cortico-cortical (frontotemporal) interactions; (iii) functional connectivity directed toward A1 from BS (feedforward)

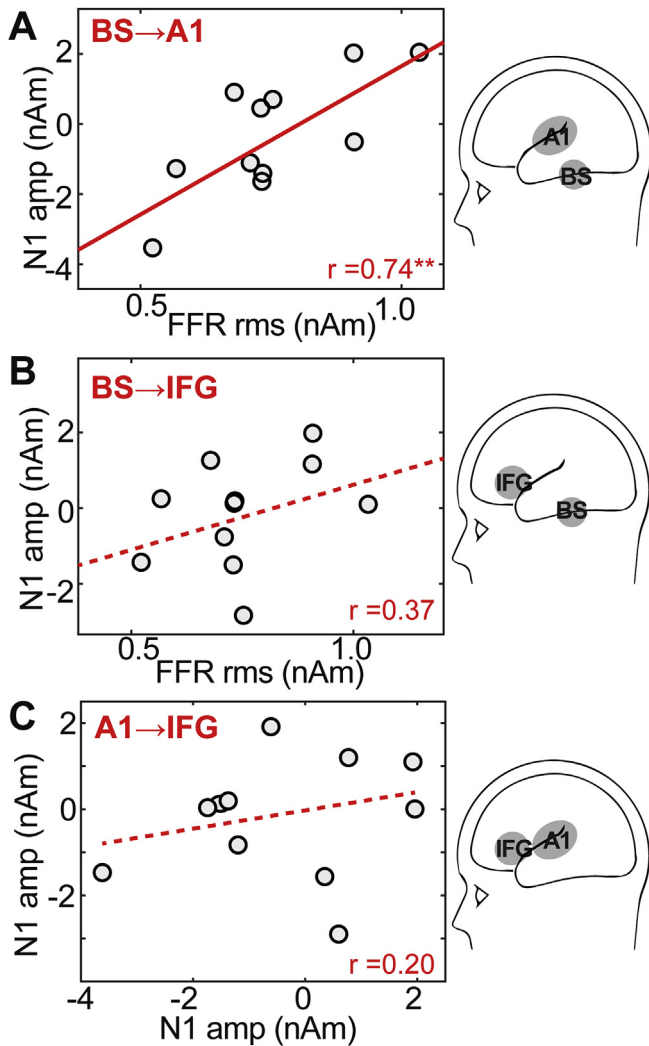
and IFG (feedback) differentiates “good” from “poor” perceivers on behavioral SIN tests.

##### 4.1. Noise and reverberation differentially impact the neural encoding of speech

Direct comparisons between acoustic stressors demonstrated that reverberation causes lesser interruption to the brain's speech representations than comparable levels of noise. These findings cannot be attributed to trivial differences in “noisiness” (i.e., acoustic figure-ground quality) as noise and reverb were matched in their severity (i.e., SNR  $\approx$  D/R). At both brainstem and cortical levels, reverberant speech produced larger evoked responses compared to clean speech, a facilitation that was most apparent in BS and IFG sources. Similarly, whole-brain analyses indicated that cortical activations were stronger for mild and moderate reverberation relative to comparable noise levels in both right A1 and left IFG (Fig. 4). The lesser impact of reverb than noise we find in electrophysiological responses is consistent with behavioral data that demonstrates human speech perception is remarkably resilient to reverberation (Bradley et al., 1999; Culling et al., 2003; Nielsen and Dau, 2010; Watkins, 2005) and the fact that listeners notice fewer signal changes in reverberation than in noise (Larsen et al., 2008; McShefferty et al., 2015; Zahorik, 2002).

Weaker cortical responses in noise could reflect neural desynchronization caused by lower inter-trial phase coherence (i.e., increased jitter) of responses in the presence of acoustic interference (Koerner and Zhang, 2015). But why is a similar desynchronization to speech not observed in reverberation? Animal recordings (Sayles and Winter 2008) and previous human FFR studies (Al Osman et al., 2017; Bidelman and Krishnan, 2010) show that speech cues are more easily maintained at the neural level in





**Fig. 6. Correlations between BS, A1, and IFG source activity during degraded speech processing.** (A) BS-FFRs are positively correlated with A1-ERP amplitudes; stronger brainstem encoding of speech is associated with larger responsivity in auditory cortex. No associations were observed between BS and IFG amplitudes (B) nor between A1 and IFG (C) suggesting speech representations become more independent at higher levels of the speech hierarchy beyond lemniscal auditory regions. Solid lines, significant correlations; dotted lines, *n.s.* correlations.  $**p < 0.01$ .

reverb compared to noise (for review, see Bidelman, 2017). Some investigators have also postulated that neurons in the brainstem inferior colliculus—the putative generator of scalp FFRs (Bidelman, 2018; Smith et al., 1975)—might perform a neural compensation that mitigates the negative effects of reverb and target signal representations (Slama and Delgutte, 2015). For a subset of collicular neurons, Slama and Delgutte (2015) found that the temporal coding of amplitude envelope was better for reverberant than for anechoic stimuli having the same modulation depth at the ear, indicating a robustness in auditory signal processing to reverb. The strong resilience of speech coding we find across brain levels may reflect population-level dynamics of these early neural compensatory mechanisms (Slama and Delgutte, 2015) that ultimately increase the tolerance of auditory processing in reverberation (Bradley et al., 1999; Culling et al., 2003; Nielsen and Dau, 2010; Watkins, 2005).

Alternatively, the resilience of neural responses to reverberation could reflect a refinement of auditory signal processing aided by corticofugal (cortico-collicular) efferent pathways. Cortical neurons can enhance the SNR of their own thalamo-collicular inputs via

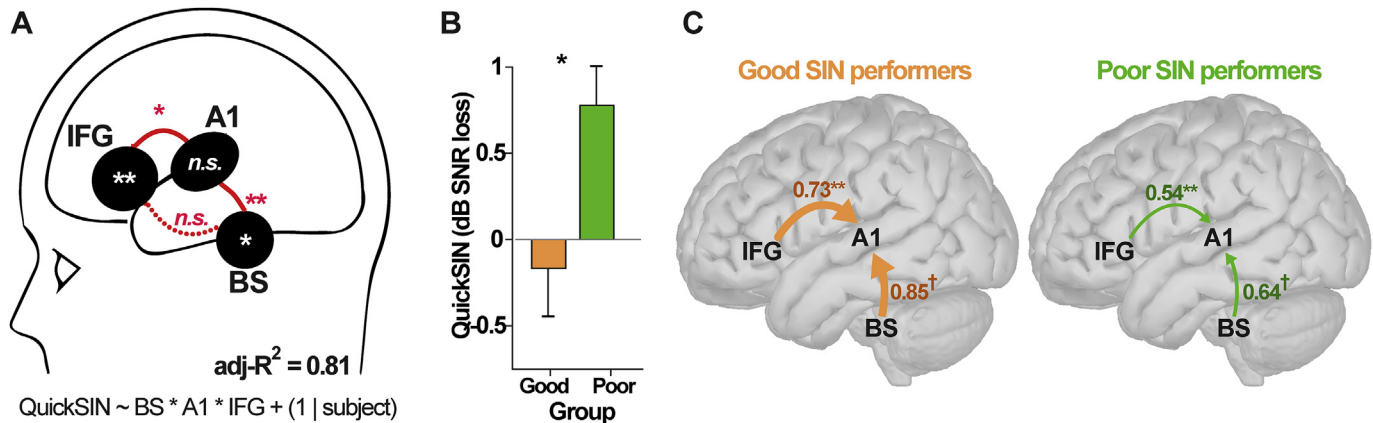
corticofugal gain but only when their receptive fields can accurately estimate task-relevant stimulus features—termed “egocentric selection” (Suga et al., 2000). However, why corticofugal effects would be specific (and/or larger) for reverberant compared to noise-degraded speech is unclear. Whether such top-down modulations can occur under strictly passive listening tasks (as used here) also remains speculative. That said, efferent modulations from the lower brainstem to the cochlea (olivocochlear bundle) that provide an “antimasking” function (Winslow and Sachs, 1987) are observed under passive scenarios and are highly predictive of SIN skills at the behavioral level (Bidelman and Bhagat, 2015; Mishra and Lutman, 2014). However, our data also show that afferent (BS→A1) rather than efferent (A1→BS) brainstem-cortical signaling differentiated “good” from “poor” listeners on the QuickSIN (Fig. 7C).

Remarkably, these brain-behavior associations were observed between passive neural encoding of relatively simple speech sounds (i.e., vCv in noise) and a more complex speech recognition task (i.e., QuickSIN), corroborating findings of recent brainstem and cortical ERP studies (Bidelman and Howell, 2016; Bidelman and Dexter, 2015; Song et al., 2011). Still, most real-world “cocktail party” scenarios require more than mere passive figure-ground (cf. target-noise) analysis as assessed here. Indeed, listeners must often parse multiple streams of competing speech information in multi-talker environments, a process itself enhanced by attention (Alain, 2007; Ding and Simon, 2012; Hill and Miller, 2010; Puvvada and Simon, 2017; Xiang et al., 2010). Future studies are needed to clarify the role of afferent/efferent interplay during speech processing and the impact of active vs. passive listening paradigms. Furthermore, while we find acoustically matched noise SNR and reverb D/R are processed differentially by the auditory system, it is still possible these interferences differ in their perceptual equivalency. Thus, an interesting extension to the present study would be to examine speech coding and brainstem-cortical connectivity for noise and reverb stimuli first titrated for perceptual difficulty. On this point, Picou et al. (2016) have recently shown that reverb induces less listening effort than noise during speech comprehension, which is consistent with our neurophysiological findings here.

#### 4.2. Hierarchical neural encoding of degraded speech

Cross-level comparisons between functional tiers revealed that noise- and reverb-related changes in neural responses were generally larger in brainstem compared to cortex. While FFRs exhibited systematic amplitude increases with increasing noise/reverb, A1 responses showed more muted changes. This suggests acoustic interferences have a differential effect on subcortical vs. cortical tiers of the auditory hierarchy during speech processing (i.e., BS > A1). To our knowledge this has not been previously reported and is only made apparent by our dual FFR-ERP paradigm.

The lesser effect of noise/reverb at the cortical compared to brainstem level might be accounted for by differences in how non-auditory, compensatory brain regions are marshalled during degraded speech processing. In cortex for instance, top-down control of auditory A1 responses from frontal regions can occur pre-attentively and independent of attention (Doeller et al., 2003). Top-down influences are also generally stronger at cortical relative to subcortical levels (Galbraith and Kane, 1993; Hillyard and Picton, 1979; Okamoto et al., 2011; Woods and Hillyard, 1978). Thus, the smaller interference-induced changes we find in A1 relative to brainstem (midbrain) could be due to stronger “top-down” influences from higher-order, non-auditory brain areas (e.g., IFG) to primary auditory cortex than those from corticofugal modulations between brainstem and A1. This notion is consistent with recent studies showing strong engagement of IFG (linguistic areas) in



**Fig. 7. Degraded speech perception is predicted by intra-regional activations and inter-regional connectivity.** (A) BS, A1, and IFG activity and their inter-regional interactions (pooled across all interference conditions) account for 81% of the variance (adj-R<sup>2</sup>) in behavioral QuickSIN scores. Cortico-collicular (BS-A1) and cortico-cortical (A1-IFG) neural interactions are also strong predictors of perception. Lines connecting ROIs show pairwise source terms that interacted in the GLME (solid lines, significant; dotted lines, n.s.) Statistical flags within ROIs reflect the main effect of that source in predicting QuickSIN scores. (B) QuickSIN scores for “good” ( $n = 6$ ) vs. “poor” ( $n = 5$ ) SIN performers (median split of scores). Despite all having normal hearing, certain listeners obtain better (~1 dB lower) SIN thresholds indicating individual differences in SIN perception. (C) Functional connectivity in the brainstem-cortical speech network differentiates “good” vs. “poor” SIN performers. Values represent the magnitude of connectivity computed via phase transfer entropy (Lobier et al., 2014). Statistical flags represent group contrasts at each connection. Feedforward (BS→A1) and feedback (IFG→A1) towards primary auditory cortex is stronger in “good” relative to “poor” perceivers. † $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

addition to A1 (primary auditory areas) during passive (Bidelman and Dexter, 2015; Bidelman and Howell, 2016) and active (Binder et al., 2004; Du et al., 2014; Scott and McGettigan, 2013) SIN processing. Indeed, our data here demonstrate that IFG is actively involved during speech processing in both distinguishing noise from reverberant speech (Fig. 4) and in predicting behavioral QuickSIN scores (Fig. 7). Our results corroborate the notion that neural activity in frontal brain regions outside conventional auditory system are robust predictors of SIN perception (Adank et al., 2012; Bidelman and Dexter, 2015; Bidelman and Howell, 2016; Binder et al., 2004; Wong et al., 2008). Nevertheless, the strong interactions we find between BS and A1 do indicate that degraded speech processing also depends on the functional interplay between subcortical and cortical hubs of the auditory hierarchy—perhaps in maintaining a veridical depiction of the speech signal during its neural ascent.

#### 4.3. Feedforward/feedback to A1 from BS and IFG is critical to degraded speech perception

In this regard, our GLME and PTE results help reveal important patterns of functional connectivity during degraded speech processing. We found intra-regional activity within auditory BS and IFG were themselves strong predictors of listeners' QuickSIN performance, in agreement with previous neuroimaging studies (cf. Bidelman and Howell, 2016; Binder et al., 2004; Coffey et al., 2017; Du et al., 2014). Similarly, BS and A1 activity was strongly correlated, suggesting that neural representations in lower auditory relays (which dominate FFRs) might feed-forward to influence sound encoding at the cortical level. These results corroborate findings of Coffey et al. (2017) who suggested that superior brainstem neural encoding of noise-degraded speech might provide better information to A1. We replicate and extend these results by showing brainstem-cortical reciprocity is also critical for speech processing in other acoustic stressors (i.e., reverb) but that behavior also depends on links between auditory and linguistic cortical regions further downstream. In addition to cortico-collicular (BS-A1) associations (Bidelman and Alain, 2015a; Coffey et al., 2017), we found that cortico-cortical (A1-IFG) interactions were a strong predictor of degraded speech perception abilities (cf. Adank et al.,

2012; Binder et al., 2004). The observation that BS and IFG responses were largely independent (Figs. 6 and 7A) further implies that speech undergoes several neural transformations from brainstem, through the auditory cortices, in route to linguistic brain regions during the analysis of speech.

Functional connectivity corroborated these regression results by demonstrating the direction of inter-regional signaling that drives degraded speech perception. In particular, we found that bottom up (BS→A1) and top down (IFG→A1) directed communication towards primary auditory cortex (A1) is critical in determining successful degraded speech perception and differentiating “good” vs. “poorer” listeners (Fig. 7). Our connectivity findings broadly agree with notions that speech is processed in a dynamic, and distributed fronto-temporal network (Bidelman and Dexter, 2015; Bidelman and Howell, 2016; Du et al., 2014; Obleser et al., 2007; Scott and McGettigan, 2013; Scott et al., 2009) whose engagement is differentially recruited depending on signal clarity, intelligibility, and linguistic experience (Adank et al., 2012; Bidelman and Dexter, 2015; Scott and McGettigan, 2013).

Our findings reveal a series of hierarchical computations are involved in degraded auditory processing, whereby neural correlates for speech are maintained in successively more abstract forms as the signal traverses the auditory system. This parallels similar hierarchical processing observed for non-speech sounds (Bidelman and Alain, 2015b). It is thought that information relayed from lower levels of the pathway are successively pruned so as to allow easier readout of signal identity in higher brain areas responsible for generating percepts (Chechik et al., 2006). Indeed, multi-unit (Perez et al., 2013) and ERP studies (e.g., Bidelman and Alain, 2015b; Bidelman et al., 2013; Bidelman et al., 2014b) directly comparing responses in brainstem and early sensory cortex help bolster the notion of a continued abstraction in the neural code (see also Sinex et al., 2005; Sinex et al., 2003). It is possible that the current data reflect a similar pruning operation where speech cues coded in lower structures (brainstem) are faithfully carried forward into cortex (A1) but are then recast upon arrival in IFG into more abstract, invariant representations that are more robust to surface-level stimulus manipulations like noise (Adank et al., 2012; Bidelman and Howell, 2016; Binder et al., 2004; Du et al., 2014).

Collectively, findings of the present study point to fundamental

differences in how noise and reverberation affect the subcortical-cortical neural encoding of speech. A differential neural coding between interferences may help account for the challenges observed by hearing impaired listeners in certain acoustic environments (but not others) as well as the unique types of perceptual confusions listeners experience in noise compared to reverberation (Nabelek and Dagenais, 1986).

## Acknowledgements

This work was supported by grants from the American Hearing Research Foundation, American Academy of Audiology (AAA) Foundation, University of Memphis Research Investment Fund (UMRIF), and NIH/NIDCD R01DC016267 awarded to G.M.B.

## References

- Adank, P., Davis, M.H., Hagoort, P., 2012. Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia* 50, 77–84.
- Afshartous, D., Preston, R.A., 2011. Key results of interaction models with centering. *J. Stat. Educ.* 19, 1–24.
- Agung, K., Purdy, S.C., McMahon, C.M., Newall, P., 2006. The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults. *J. Am. Acad. Audiol.* 17, 559–572.
- Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.
- Al Osman, R., Dajani, H.R., Giguère, C., 2017. Self-masking and overlap-masking from reverberation using the speech-evoked auditory brainstem response. *J. Acoust. Soc. Am.* 142, EL555–EL560.
- Alain, C., 2007. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236.
- Alain, C., McDonald, K., Van Roon, P., 2012. Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. *Hear. Res.* 283, 126–135.
- Alain, C., Roye, A., Salloum, C., 2014. Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Front. Syst. Neurosci.* 8, 8.
- Alain, C., Arsenault, J.S., Garami, L., Bidelman, G.M., Snyder, J.S., 2017. Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Sci. Rep.* 7, 1–11.
- Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010. Neural timing is linked to speech perception in noise. *J. Neurosci.* 30, 4922–4926.
- Barnett, L., Barrett, A.B., Seth, A.K., 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* 103, 238701.
- Bellier, L., Bouchet, P., Jeanvoine, A., Valentin, O., Thai-Van, H., Caclin, A., 2015. Topographic recordings of auditory evoked potentials to speech: subcortical and cortical responses. *Psychophysiology* 52, 594–599.
- Berg, P., Scherg, M., 1994. A fast method for forward computation of multiple-shell spherical head models. *Electroencephalogr. Clin. Neurophysiol.* 90, 58–64.
- Bidelman, G.M., 2015. Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J. Neurosci. Meth.* 241, 94–100.
- Bidelman, G.M., 2016. Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *J. Acoust. Soc. Am.* 140, EL358–EL363.
- Bidelman, G.M., 2017. Communicating in challenging environments: noise and reverberation. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), *Springer Handbook of Auditory Research: The Frequency-following Response: a Window into Human Communication*. Springer Nature, New York, N.Y.
- Bidelman, G.M., 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69.
- Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125.
- Bidelman, G.M., Bhagat, S.P., 2015. Right ear advantage drives the link between olivocochlear efferent “antimasking” and speech-in-noise listening benefits. *Neuroreport* 26, 483–487.
- Bidelman, G.M., Alain, C., 2015a. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.* 35, 1240–1249.
- Bidelman, G.M., Dexter, L., 2015. Bilinguals at the “cocktail party”: dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners’ speech-in-noise recognition deficits. *Brain Lang.* 143, 32–41.
- Bidelman, G.M., Alain, C., 2015b. Hierarchical neurocomputations underlying concurrent sound segregation: connecting periphery to percept. *Neuropsychologia* 68, 38–50.
- Bidelman, G.M., Howell, M., 2016. Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. *Neuroimage* 124, 581–590.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cognit. Neurosci.* 23, 425–434.
- Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212.
- Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014a. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673.
- Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014b. Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35, 2526–2540.
- Bidelman, G.M., Lowther, J.E., Tak, S.H., Alain, C., 2017. Mild cognitive impairment is characterized by deficient hierarchical speech coding between auditory brainstem and cortex. *J. Neurosci.* 37, 3610–3620.
- Billings, C.J., Tremblay, K.L., Stecker, G.C., Tolin, W.M., 2009. Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear. Res.* 254, 15–24.
- Billings, C.J., Bennett, K.O., Molis, M.R., Leek, M.R., 2010. Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear.* 32, 53–60.
- Billings, C.J., McMillan, G.P., Penman, T.M., Gille, S.M., 2013. Predicting perception in noise using cortical auditory evoked potentials. *J. Assoc. Res. Oto* 14, 891–903.
- Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301.
- Bradley, J.S., Reich, R.D., Norcross, S.G., 1999. On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *J. Acoust. Soc. Am.* 106, 1820–1828.
- Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear.* 33, 144–150.
- Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432.
- Chechik, G., Anderson, M.J., Bar-Yosef, O., Young, E.D., Tishby, N., Nelken, I., 2006. Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51, 359–368.
- Coffey, E.B.J., Chpeusiuk, A.M.P., Herholz, S.C., Baillet, S., Zatorre, R.J., 2017. Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Front. Neurosci.* 11.
- Culling, J.F., Hodder, K.I., Toh, C.Y., 2003. Effects of reverberation on perceptual segregation of competing voices. *J. Acoust. Soc. Am.* 114, 2871–2876.
- Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859.
- Doeller, C.F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., Schröger, E., 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20, 1270–1282.
- Du, Y., Buchsbaum, B.R., Grady, C.L., Alain, C., 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1–6.
- Fujihira, H., Shiraishi, K., 2015. Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clin. Neurophysiol.* 126, 96–102.
- Galbraith, G.C., Kane, J.M., 1993. Brainstem frequency-following responses and cortical event-related potentials during attention. *Percept. Mot. Skills* 76, 1231–1241.
- Gao, E., Suga, N., 1998. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12663–12670.
- Goldenholz, D.M., Ahlfors, S.P., Hämäläinen, M.S., Sharon, D., Ishitobi, M., Vaina, L.M., Stufflebeam, S.M., 2009. Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Hum. Brain Mapp.* 30, 1077–1086.
- Helfer, K., Wilber, L., 1990. Hearing loss, aging, and speech perception in reverberation and in noise. *J. Speech Hear. Res.* 33, 149–155.
- Herdman, A.T., Lins, O., van Roon, P., Stapells, D.R., Scherg, M., Picton, T., 2002. Intracerebral sources of human auditory steady-state responses. *Brain Topogr.* 15, 69–86.
- Hill, K.T., Miller, L.M., 2010. Auditory attentional control and selection during cocktail party listening. *Cerebr. Cortex* 20, 583–590.
- Hillebrand, A., Tewarie, P., van Dellen, E., Yu, M., Carbo, E.W.S., Douw, L., Gouw, A.A., van Straaten, E.C.W., Stam, C.J., 2016. Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3867–3872.
- Hillyard, S.A., Picton, T.W., 1979. Event-related brain potentials and selective information processing in man. In: Desmedt, J.E. (Ed.), *Progress in Clinical Neurophysiology*. Karger, Basel, pp. 1–52.
- Humes, L.E., Roberts, L., 1990. Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility. *J. Speech Hear. Res.* 33, 726–735.
- Iordanov, T., Hoechstetter, K., Berg, P., Paul-Jordanov, I., Scherg, M., 2014. CLARA: Classical LORETA Analysis Recursively Applied, OHBM 2014.
- Iordanov, T., Bornfleth, H., Hoechstetter, K., Lanfer, B., Scherg, M., 2016. Performance of Cortical LORETA and Cortical CLARA Applied to MEG Data, Biomag 2016.
- Killion, M., Niquette, P., 2000. What can the pure-tone audiogram tell us about a patient’s snr loss? *Hear. J.* 53, 46–53.
- Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., Banerjee, S., 2004.

- Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116, 2395–2405.
- Kinsler, L.E., Frey, A.R., Coppens, A.B., Sanders, J.V., 2000. *Fundamentals of Acoustics*. John Wiley & Sons, Inc., New York, NY.
- Koerner, T.K., Zhang, Y., 2015. Effects of background noise on inter-trial phase coherence and auditory N1–P2 responses to speech stimuli. *Hear. Res.* 328, 113–119.
- Kraus, N., Cheour, M., 2000. Speech sound representation in the brain. *Audiol. Neuro. Otol.* 5, 140–150.
- Kraus, N., White-Schwoch, T., 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cognit. Sci.* 19, 642–654.
- Krishnan, A., 2007. Human frequency following response. In: Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins, Baltimore, pp. 313–335.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Res.* 1313, 124–133.
- Krishnan, A., Bidelman, G.M., Smalt, C.J., Ananthakrishnan, S., Gandour, J.T., 2012. Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia* 50, 2849–2859.
- Krishnaswamy, P., Obregon-Henafo, G., Ahveninen, J., Khan, S., Babadi, B., Iglesias, J.E., Hamalainen, M.S., Purdon, P.L., 2017. Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10465–E10474.
- Larsen, E., Iyer, N., Lansing, C.R., Feng, A.S., 2008. On the minimum audible difference in direct-to-reverberant energy ratio. *J. Acoust. Soc. Am.* 124, 450–461.
- Lobier, M., Siebenhuhner, F., Palva, S., Palva, J.M., 2014. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage* 85 (2), 853–872.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Meth.* 164, 177–190.
- Mazziotta, J.C., Toga, A.W., Evans, A., Lancaster, J.L., Fox, P.T., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* 2, 89–101.
- McShefferty, D., Whitmer, W.M., Akeroyd, M.A., 2015. The just-noticeable difference in speech-to-noise ratio. *Trends in Hearing* 19, 1–9.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R., 2004. EEG source imaging. *Clin. Neurophysiol.* 115, 2195–2222.
- Middelweerd, M.J., Festen, J.M., Plomp, R., 1990. Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology* 29, 1–7.
- Mishra, S.K., Lutman, M.E., 2014. Top-down influences of the medial olivocochlear efferent system in speech perception in noise. *PLoS One* 9, e85756.
- Muscaccia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42.
- Nabelek, A.K., Dagenais, P.A., 1986. Vowel errors in noise and in reverberation by hearing-impaired listeners. *J. Acoust. Soc. Am.* 80, 741–748.
- Nabelek, A.K., Letowski, T.R., Tucker, F.M., 1989. Reverberant overlap- and self-masking in consonant identification. *J. Acoust. Soc. Am.* 86, 1259–1265.
- Nielsen, J.B., Dau, T., 2010. Revisiting perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* 128, 3088–3094.
- Obleser, J., Wise, R.J., Dresner, M.A., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289.
- Okamoto, H., Stracke, H., Bermudez, P., Pantev, C., 2011. Sound processing hierarchy within human auditory cortex. *J. Cognit. Neurosci.* 23, 1855–1863.
- Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiology* 112, 713–719.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9.
- Otnes, R., Enochson, L., 1972. *Digital Time Series Analysis*. Wiley.
- Parbery-Clark, A., Skoe, E., Kraus, N., 2009a. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009b. Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- Parbery-Clark, A., Marmel, F., Bair, J., Kraus, N., 2011. What subcortical-cortical relationships tell us about processing speech in noise. *Eur. J. Neurosci.* 33, 549–557.
- Pascual-Marqui, R.D., Esslen, M., Kochi, K., Lehmann, D., 2002. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find. Exp. Clin. Pharmacol.* 24 (Suppl. C), 91–95.
- Perez, C.A., Engineer, C.T., Jakkamsetti, V., Carraway, R.S., Perry, M.S., Kilgard, M.P., 2013. Different timescales for the neural coding of consonant and vowel sounds. *Cerebr. Cortex* 23, 670–683.
- Picou, E.M., Gordon, J., Ricketts, T.A., 2016. The effects of noise and reverberation on listening effort for adults with normal hearing. *Ear Hear.* 37, 1–13.
- Picton, T.W., van Roon, P., Armiljo, M.L., Berg, P., Ille, N., Scherg, M., 2000. The correction of ocular artifacts: a topographic perspective. *Clin. Neurophysiol.* 111, 53–65.
- Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., Trujillo, N.J., 1999. Intracerebral sources of human auditory-evoked potentials. *Audiol. Neuro. Otol.* 4, 64–79.
- Puvvada, K.C., Simon, J.Z., 2017. Cortical representations of speech in a multitalker auditory scene. *J. Neurosci.* 37, 9189–9196.
- Richards, J.E., Sanchez, C., Phillips-Meek, M., Xie, W., 2016. A database of age-appropriate average MRI templates. *Neuroimage* 124, 1254–1259.
- Ross, B., Tremblay, K., 2009. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hear. Res.* 248, 48–59.
- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the bio-magnetic inverse problem. *Phys. Med. Biol.* 32, 11–22.
- Sayles, M., Winter, I.M., 2008. Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58, 789–801.
- Scherg, M., Ebersole, J.S., 1994. Brain source imaging of focal and multifocal epileptiform EEG activity. *Neurophysiol. Clin.* 24, 51–60.
- Scherg, M., Ille, N., Bornfleth, H., Berg, P., 2002. Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. *J. Clin. Neurophysiol.* 19, 91–112.
- Schonwiesner, M., Krumbholz, K., Rubsamen, R., Fink, G.R., von Cramon, D.Y., 2007. Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. *Cerebr. Cortex* 17, 492–499.
- Scott, S.K., McGettigan, C., 2013. The neural processing of masked speech. *Hear. Res.* 303, 58–66.
- Scott, S.K., Rosen, S., Beaman, C.P., Davis, J.P., Wise, R.J., 2009. The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *J. Acoust. Soc. Am.* 125, 1737–1743.
- Sharma, A., Dorman, M.F., 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083.
- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R.J., Naatanen, R., 1998. Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neurosci. Lett.* 251, 141–144.
- Sinex, D.G., Li, H., Velenovsky, D.S., 2005. Prevalence of stereotypical responses to mistuned complex tones in the inferior colliculus. *J. Neurophysiol.* 94, 3523–3537.
- Sinex, D.G., Guzik, H., Li, H., Henderson, S.J., 2003. Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res.* 182, 130–139.
- Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324.
- Slama, M.C.C., Delgutte, B., 2015. Neural coding of sound envelope in reverberant environments. *J. Neurosci.* 35, 4452–4468.
- Slugocki, C., Bosnyak, D., Trainor, L., 2017. Simultaneously-evoked auditory potentials (SEAP): a new method for concurrent measurement of cortical and subcortical auditory-evoked activity. *Hear. Res.* 345, 30–42.
- Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr. Clin. Neurophysiol.* 39, 465–472.
- Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011. Perception of speech in noise: neural correlates. *J. Cognit. Neurosci.* 23, 2268–2279.
- Suga, N., Xiao, Z., Ma, X., Ji, W., 2002. Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18.
- Suga, N., Gao, E., Zhang, Y., Ma, X., Olsen, J.F., 2000. The corticofugal system for hearing: recent progress. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11807–11814.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., Otis, B., 2001. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear.* 22, 79–90.
- Vicente, R., Wibral, M., Lindner, M., Pipa, G., 2011. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* 30, 45–67.
- von Békésy, G., 1938. Über die Entstehung der Entfernungsempfindung beim Hören (On the origin of distance perception in hearing). *Akust. Zh.* 3, 21–31.
- Wang, D.L., Brown, G.J., 2006. *Computational Auditory Scene Analysis*. Wiley, Hoboken, NJ.
- Watkins, A.J., 2005. Perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* 118, 249–262.
- White-Schwoch, T., Kraus, N., 2017. The janus face of auditory learning: how life in sound shapes everyday communication. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), *Springer Handbook of Auditory Research: the Frequency-following Response: a Window into Human Communication*. Springer Nature, New York, N.Y., pp. 121–158.
- Winslow, R.L., Sachs, M.B., 1987. Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. *J. Neurophysiol.* 57, 1002–1021.
- Wong, P.C.M., Uppunda, A.K., Parrish, T.B., Dhar, S., 2008. Cortical mechanisms of speech perception in noise. *J. Speech Lang. Hear. Res.* 51, 1026–1041.
- Woods, D.L., Hillyard, S.A., 1978. Attention at the cocktail party: brainstem evoked responses reveal no peripheral gating. In: Otto, D.A. (Ed.), *Multidisciplinary Perspectives in Event-related Brain Potential Research (EPA 600/9-77-043)*. U.S. Government Printing Office, Washington, DC, pp. 230–233.
- Xiang, J., Simon, J., Elhilali, M., 2010. Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. *J. Neurosci.* 30, 12084–12093.
- Zahorik, P., 2002. Direct-to-reverberant energy ratio sensitivity. *J. Acoust. Soc. Am.* 112, 2110–2117.
- Zendel, B.R., Alain, C., 2009. Concurrent sound segregation is enhanced in musicians. *J. Cognit. Neurosci.* 21, 1488–1498.
- Zendel, B.R., Alain, C., 2014. Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiol. Aging* 35, 55–63.