Differential responses to spectrally degraded speech within human auditory cortex: An intracranial electrophysiology study

Kirill V. Nourski[#1,2], **Mitchell Steinschneider**[#3], **Ariane E. Rhone**[1], **Christopher K. Kovach**[1], **Hiroto Kawasaki**[1], and **Matthew A. Howard III**[1,2,4]

[1]Department of Neurosurgery, The University of Iowa, Iowa City, IA, USA

[2]Iowa Neuroscience Institute, The University of Iowa, Iowa City, IA, USA

[3]Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA

[4]Pappajohn Biomedical Institute, The University of Iowa, Iowa City, IA, USA

[#] These authors contributed equally to this work.

Abstract

Understanding cortical processing of spectrally degraded speech in normal-hearing subjects may provide insights into how sound information is processed by cochlear implant (CI) users. This study investigated electrocorticographic (ECoG) responses to noise-vocoded speech and related these responses to behavioral performance in a phonemic identification task. Subjects were neurosurgical patients undergoing chronic invasive monitoring for medically refractory epilepsy. Stimuli were utterances /aba/ and /ada/, spectrally degraded using a noise vocoder (1-4 bands). ECoG responses were obtained from Heschl's gyrus (HG) and superior temporal gyrus (STG), and were examined within the high gamma frequency range (70-150 Hz). All subjects performed at chance accuracy with speech degraded to 1 and 2 spectral bands, and at or near ceiling for clear speech. Inter-subject variability was observed in the 3 and 4-band conditions. High gamma responses in posteromedial HG (auditory core cortex) were similar for all vocoded conditions and clear speech. A progressive preference for clear speech emerged in anterolateral segments of HG, regardless of behavioral performance. On the lateral STG, responses to all vocoded stimuli were larger in subjects with better task performance. In contrast, both behavioral and neural responses to clear speech were comparable across subjects regardless of their ability to identify degraded stimuli. Findings highlight differences in representation of spectrally degraded speech across cortical areas and their relationship to perception. The results are in agreement with prior non-invasive results. The data provide insight into the neural mechanisms associated with variability in perception of degraded speech and potentially into sources of such variability in CI users.

Corresponding author: Kirill V. Nourski, MD, PhD, Department of Neurosurgery, University of Iowa Hospitals and Clinics, 200 Hawkins Dr. 1815 JCP, Iowa City, IA 52242 USA, Phone: +1 (319)353-4494, kirill-nourski@uiowa.edu.

Keywords

1   Introduction

A valuable experimental approach used to gain indirect insights into how sound information is processed by cochlear implant (CI) users involves presenting spectrally degraded speech to normal-hearing listeners. This method is implemented with a noise vocoder, which is designed to approximate the output of the CI to the auditory system. It does so by limiting the spectral content of sounds to a discrete number of frequency bands using noise carriers (Merzenich, 1983). This process preserves the gross temporal envelope of the stimulus while degrading spectral complexity and thus impairing intelligibility (Shannon et al., 1995; Dorman et al., 1997). Information gained by studying cortical processing of noise-vocoded speech in normal-hearing listeners has been used to assist in the optimization of CI designs and post-implantation rehabilitation strategies (Harnsberger et al., 2001; Li and Fu, 2007; Smalt et al., 2013).

There is considerable variability in speech perception outcomes in CI users (e.g. Blamey et al., 1996, 2013; Anderson et al., 2017). Beyond obvious variables related to the specific design of the CI and surgical placement of the electrode array within the cochlea, multiple other factors have been identified, including duration and level of hearing loss prior to implantation (Blamey et al., 1996, 2013; Lazard et al., 2012; Holden et al., 2013; Beyea et al., 2016). However, a significant amount of variability in post-implantation speech perception remains even after these and other factors are accounted for (Lazard et al., 2012; Moberly et al., 2016). It has been suggested that much of this remaining CI performance variability is based upon differences in activation within auditory cortex (Giraud & Lee, 2007; Finke et al., 2016; Anderson et al., 2017). Supporting evidence has been obtained from multiple neuroimaging (positron emission tomography) studies of CI users, which have demonstrated a correlation between strength of activation within non-primary auditory cortex and performance on speech recognition tests (e.g. Fujiki et al., 1999; Green et al., 2005; Mortensen et al., 2006; Petersen et al., 2013).

Patterns of auditory cortical activation by noise-vocoded speech in normal-hearing listeners have been examined in multiple non-invasive studies. Disagreement remains concerning the degree to which intelligible speech is processed bilaterally or is left-lateralized, and the relative contributions of anterior and posterior portions of superior temporal cortex (Scott et al., 2000, 2006; Okada et al., 2010; Evans et al., 2014). In the current study, direct electrocorticographic (ECoG) recordings from auditory cortex were used to examine high gamma band activity elicited by noise-vocoded speech in normal-hearing listeners. The primary goal was to characterize the neural signatures underlying auditory cortical processing of noise-vocoded speech and relate them to the variability in perception of these degraded signals. The secondary goal was to determine whether specific regions of auditory cortex processed noise-vocoded speech in different ways, and if so, how these responses compared to neural responses elicited by clear speech. As part of this goal, response patterns

from both hemispheres were also compared. Neural activity in Heschl's gyrus (HG) and on the lateral surface of the superior temporal gyrus (STG) was examined. These regions are envisioned to represent portions of auditory core, belt, and parabelt cortex (e.g. Sweet et al., 2005; Moerel et al., 2014; Hackett, 2015; Leaver & Rauschecker, 2016), and thus offer the opportunity to examine transformations of neural activity across the hierarchically organized auditory cortex. Finally, it was hoped that these recordings would provide insights into the relationship between strength of cortical activation and CI performance.

2 Material and methods

2.1 Subjects

Experimental subjects were ten neurosurgical patient volunteers (7 male, 3 female, age 20–51 years old, median age 31.5 years old) with pharmacoresistant epilepsy who were undergoing chronic invasive ECoG monitoring to identify potentially resectable seizure foci. All subjects had left-hemisphere language dominance, as determined by the Wada test. Research protocols were approved by the University of Iowa Institutional Review Board and by the National Institutes of Health. Written informed consent was obtained from each subject. Participation in the research protocol did not interfere with acquisition of clinically required data. Subjects could rescind consent at any time without interrupting their clinical evaluation. The patients were typically weaned from their antiepileptic medications during the monitoring period at the discretion of their treating neurologist. Experimental sessions were suspended for at least three hours if a seizure occurred, and the patient had to be alert and willing to participate for the research activities to resume.

The left hemisphere was implanted with electrodes in five subjects, and the right hemisphere was implanted in four subjects; the remaining subject (B335) had bihemispheric electrode coverage. The side of implantation is indicated by the prefix of the subject code: L for left, R for right, B for bilateral. Intracranial recordings revealed that the auditory and auditory-related cortical areas from which experimental data were obtained were not epileptic foci in any of the subjects.

All subjects except one (L275) were native English speakers. Subject L275 was a native Bosnian speaker who learned German at the age of 10 and English at the age of 17. All subjects except two (L282 and L307) had pure-tone thresholds within 20 dB hearing level (HL) between 250 Hz and 4 kHz. Subject L282 had a moderate (55 dB) 4 kHz notch in both ears. Subject L307 had a mild (40 dB) 4 kHz notch in the right ear. Word recognition scores, as evaluated by spondee words, were 100/100% (right/left ear) in all tested subjects except one (R316), who scored 96/100%. Speech reception thresholds were within 15 dB HL in all tested subjects, including the two with tone audiometry thresholds outside the 20 dB HL criterion.

2.2 Stimuli and procedure

Experimental stimuli were the utterances /aba/ and /ada/, spoken by a male talker (Tyler et al., 1989). The stimuli were spectrally degraded using a noise vocoder with 1, 2, 3 or 4 frequency bands following the approach of Shannon et al. (1995) (Supplementary Fig. 1).

Vocoded speech sounds were generated using a modification of Dr. Chris Darwin's Shannon script (http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/Shannon) implemented in a Praat v.5.2.03 environment (Boersma, 2001). Filter cutoff frequencies were 1500 Hz for the 2-band vocoder, 800 and 1500 Hz for the 3-band vocoder, and 800, 1500 and 2500 Hz for the 4-band vocoder.

Temporal envelopes were low-pass filtered at 50 Hz, and all stimuli, including the original (clear) utterances, were low-pass filtered at 4 kHz. The duration of /aba/ and /ada/ was 497 and 467 ms, respectively. The utterance /aba/ had steady-state vowel fundamental frequencies (F_0) of 136 and 137 Hz for the first and the second /a/, respectively, while /ada/ had steady-state vowel fundamental frequencies of 130 and 112 Hz for the first and the second /a/, respectively (see Supplementary Fig. 1, right column).

Experiments were carried out in a dedicated, electrically-shielded suite in The University of Iowa Clinical Research Unit. The room was quiet, with lights dimmed. Subjects were awake and reclining in a hospital bed or an armchair. The stimuli were delivered at a comfortable level (typically 60-65 dB SPL) via insert earphones (Etymotic ER4B, Etymotic Research, Elk Grove Village, IL, USA) coupled to custom-fit earmolds. Stimulus delivery was controlled using Presentation software (Version 16.5; Neurobehavioral Systems).

Stimuli (40 repetitions of each of the 10 stimuli; see Supplementary Fig. 1) were presented in random order in a one-interval yes-no discrimination task (Macmillan & Creelman, 2004). Subjects were instructed to report whether they heard an /aba/ or an /ada/ on each trial by pressing one of the two trigger buttons on a Microsoft Sidewinder video game controller: left for /aba/ and right for /ada/. The two choices were shown to the subject on a computer screen, and 250 ms following the subject's button press the correct answer was highlighted for 250 ms to provide real-time feedback on the task performance. The next trial was presented following a delay of 750-760 ms. The complete task took between 15 and 30 minutes to complete (median duration 20 minutes). Longer task times lead to significant fatigue in these patient/subjects, thus limiting the number of vocoded speech conditions that could reasonably be presented in this experimental paradigm.

2.3 Recording

ECoG recordings were made from depth electrodes implanted in HG and subdural electrodes overlying the lateral surface of the STG. Electrode implantation, recording and ECoG data analysis have been previously described in detail (Howard et al., 1996, 2000; Reddy et al., 2010; Nourski and Howard, 2015). All electrodes were placed solely on the basis of clinical requirements to identify seizure foci (Nagahama et al., 2018). In all subjects the temporal neocortex was suspected to be involved in either generation or propagation of seizures. To that end, lateral and ventral temporal cortex was sampled with subdural electrodes and dorsal and medial temporal neocortex was sampled with depth electrodes that included the HG electrode.

Electrode arrays were manufactured by Ad-Tech Medical (Racine, WI). Depth electrode arrays (4-8 macro contacts, spaced 5-10 mm apart) targeting HG were stereotactically implanted in each subject along the anterolateral-to-posteromedial axis of the gyrus. Grid

arrays consisted of platinum-iridium disc electrodes (2.3 mm exposed diameter, 5-10 mm inter-electrode distance) embedded in a silicon membrane. A subgaleal electrode was used as a reference. ECoG data acquisition was controlled by a TDT RZ2 real-time processor (Tucker-Davis Technologies, Alachua, FL). Collected ECoG data were amplified, filtered (0.7–800 Hz bandpass, 12 dB/octave rolloff), digitized at a sampling rate of 2034.5 Hz, and stored for subsequent offline analysis.

2.4 Analysis

Subjects' performance on the behavioral task was characterized in terms of accuracy (% correct), sensitivity (d') and reaction times (RT). For each of the five stimulus conditions (1-, 2-, 3-, 4-band and clear), statistical significance of across-subject median task accuracy relative to chance performance was established using one-tailed Wilcoxon signed rank tests. Differences in task accuracy and RT between adjacent conditions were established using two-tailed Wilcoxon signed-rank tests. Correlation between median RTs and task accuracy across subjects was measured using Spearman rank correlation tests. Subjects were rank-ordered according to their average accuracy in the 3- and 4-band conditions to examine the relationship between neural activity and task performance. This level of spectral degradation approximates a midpoint between chance and ceiling performance for identification of the two stop consonants varying in their place of articulation (Shannon et al., 1995; McGettigan et al., 2014). Good and poor task performance was determined by each subject's average accuracy in the 3- and 4- band conditions, with good performance defined as average accuracy >67.5% (i.e. above the across-subject median; see Results) and $d' > 1$.

To investigate possible relationship between the subjects' hearing and task performance, pure-tone average values were computed as average audiogram thresholds for tone frequencies of 0.5, 1 and 2 kHz. Better-ear pure-tone averages in each subject were then correlated with their average accuracy in 3- and 4-band conditions using non-parametric statistics (Spearman's rank-order correlation).

Reconstruction of the anatomical locations of the implanted electrodes and their mapping onto a standardized set of coordinates across subjects was performed using FreeSurfer image analysis suite (Version 5.3; Martinos Center for Biomedical Imaging, Harvard, MA) and in-house software (see Nourski et al., 2014, for details). In brief, subjects underwent whole-brain high-resolution T1-weighted structural magnetic resonance imaging (MRI) scans (resolution and slice thickness   1.0 mm) before electrode implantation. After electrode implantation, subjects underwent MRI and thin-slice volumetric computerized tomography (CT) (resolution and slice thickness   1.0 mm) scans. Contact locations of the HG depth electrodes and subdural grid electrodes were first extracted from post-implantation MRI and CT scans, respectively. These were then projected onto preoperative MRI scans using non-linear threedimensional thin-plate spline morphing, aided by intraoperative photographs. For group analyses, standard Montreal Neurological Institute (MNI) coordinates were obtained for each contact using linear coregistration to the MNI152 T1 average brain, as implemented in FMRIB Software library (Version 5.0; FMRIB Analysis Group, Oxford, UK). Left hemisphere MNI x-axis coordinates (x_{MNI}) were then multiplied by (-1) to map them onto the right-hemisphere common space.

Recording sites were included in analyses based on their anatomical location (i.e., implanted in the HG or overlying the lateral surface of the STG) as determined by the localization of each electrode in the pre-implantation MR for each subject individually, rather than based on common MNI coordinates. Based on these criteria, a total of 73 HG recording sites and 208 sites on lateral STG from the 10 subjects were examined. In subject R320, the depth electrode trajectory was localized to the gray matter of the anterior transverse sulcus. These contacts were included in analysis with sites in the crest of HG, as core auditory cortex extends onto the gray matter within the sulci surrounding HG (e.g. Rademacher et al., 2001; Da Costa et al., 2011).

To facilitate a straightforward interpretation of depth electrode locations in terms of their orientation along HG, the MNI coordinates were rotated along the anatomical HG axis in the $x_{MNI}y_{MNI}$ plane, as previously described (Nourski et al., 2014). The coordinates in this plane were first centered by subtracting the across-electrode mean from each individual coordinate. Next, the best fit linear regression line was computed between x_{MNI} and y_{MNI}. The angle of rotation, θ, was computed from the slope of that line, and each set of coordinates was rotated by θ. This process resulted in the new x_θ coordinate which defined the position along the long axis of HG with coordinate values increasing from posteromedial to anterolateral. Likewise, to describe the locations of electrodes overlying the lateral STG along the length of the gyrus, their MNI coordinates were centered and rotated along anatomical STG axis in the $y_{MNI}z_{MNI}$ plane, yielding new y_θ and z_θ coordinates that characterized each electrode's location relative to the posterior-anterior and dorsal-ventral dimension of the gyrus, respectively. HG and lateral STG sites were then divided into equal-width groups (three for each gyrus) along the long axis of each gyrus. This procedure permitted a finer-grained analysis of activity profiles in auditory cortex and is based on previous studies showing gradients of sound-evoked response properties along these gyri (Nourski et al., 2014, 2017). These six groups served as our regions of interest (ROIs). Finally, locations of electrodes overlying the lateral STG relative to the ventral-dorsal dimension of the gyrus (i.e. from the Sylvian fissure and the superior temporal sulcus) were described based on their z_θ coordinates, with negative and positive values corresponding to ventral and dorsal aspects of the lateral STG, respectively.

Auditory cortical activity was measured and characterized as event-related band power (ERBP). Trials with any voltage deflections greater than five standard deviations from the mean calculated over the entire duration of the recording were assumed to be artifact and thus excluded from subsequent analyses. Time-frequency analysis was carried out using a demodulated band transform method (Kovach & Gander, 2016), with software available at https://github.com/ckovach/DBT. In brief, this approach obtains the discrete Fourier transform (DFT) of the entire signal, segments the DFT into short overlapping intervals, windows each segment with a cosine window and applies the inverse DFT to each segment. ECoG signal power was computed within overlapping frequency windows of variable (1-20 Hz) bandwidth for theta (center frequencies 4-8 Hz, 1 Hz step), alpha (8-14 Hz, 2 Hz step), beta (14-30 Hz, 4 Hz step), gamma (30-70 Hz, 10 Hz step) and high gamma (70-150 Hz, 20 Hz step) ECoG bands. For each center frequency, the squared modulus of the resulting complex signal was log-transformed, segmented into single trial epochs, normalized by subtracting the mean log power within a reference interval (100-200 ms before stimulus

onset in each trial), and averaged over trials to obtain ERBP for each center frequency. Average ERBP was computed for each level of spectral degradation in the stimulus, pooling over /aba/ and /ada/ conditions.

Quantitative analyses focused on high gamma ERBP, which is closely related to unit activity and thus largely reflects the output signal from the recorded region (Mukamel et al., 2005; Nir et al., 2007; Ray et al., 2008; Steinschneider et al., 2008; Whittingstall & Logothetis, 2009). High gamma ERBP was calculated by averaging power envelopes for center frequencies between 70 and 150 Hz. Mean high gamma ERBP values were computed within 50-550 ms windows (relative to stimulus onsets) and averaged across trials for each level of spectral degradation and for each recording site. This time interval corresponded to the approximate duration of the stimuli with a 50-ms delay to account for the approximate latency of auditory cortical high gamma responses (e.g. Nourski et al., 2014). The spatial distribution of high gamma activity simultaneously acquired across the entire subdural electrode grid was examined in subject L275. To that end, mean high gamma ERBP values were smoothed over the 8×12 contact grid using triangle-based cubic interpolation with an upsampling factor of 16.

Differences in across-electrode median high gamma ERBP values between adjacent conditions (e.g. 1- vs. 2-band, 2- vs. 3-band, etc.) were tested in each ROI for significance using Wilcoxon signed rank tests with false discovery rate (FDR) correction for multiple comparisons (Benjamini & Hochberg, 1995). At the single-electrode level, significance of high gamma responses for the three levels of spectral degradation (1- & 2-band combined, 3- & 4-band combined, and clear speech) was established using one-sample one-tailed t-tests (significance threshold $p = 0.05$, FDR-corrected). Comparison of high gamma ERBP recorded from the STG in subjects who exhibited good versus poor performance on the task was performed using Wilcoxon rank sum test for 1- & 2-band combined, 3- & 4-band combined, and clear speech.

3 Results

3.1 Behavioral performance in the one-interval yes-no discrimination task

All subjects performed at chance level when attempting to identify speech utterances spectrally degraded with 1 and 2 bands ($p = 0.539$, $W = 17.5$, and $p = 0.450$, $W = 29$, respectively, one-tailed Wilcoxon signed rank tests) (Fig. 1, top panel). There also was no systematic difference between these two unintelligible conditions ($p = 0.906$, $W = 26$; two-tailed Wilcoxon signed rank test). At the other extreme, all subjects were able to identify clear utterances, performing above chance ($p = 9.77 \times 10^{-4}$, $W = 55$, one-tailed Wilcoxon signed rank test), and at or near ceiling level.

In contrast to the uniformity of behavioral performance at the extremes of intelligibility, marked variability was observed at intermediate degrees of spectral degradation (3 and 4 bands). This variability parallels that seen in previous work (Shannon et al., 1995; McGettigan et al., 2014). In these two intermediate stimulus conditions, five subjects (L237, L307, R288, L275, R320) performed above chance with accuracy >67.5% and d' >1 (Fig. 1, top and middle panels). Even though subject B335 had a gradual improvement in accuracy

between the 3- and 4-band conditions, d' was still <1 in the latter condition. Subject R316 was atypical from all other subjects, exhibiting above-chance accuracy and sensitivity at 3-band, yet performing at chance at 4-band condition. Subject R263 exhibited an average accuracy of 66.3% for the 3- and 4-band condition, yet this subject's performance was characterized by relatively low sensitivity ($d = 0.842$ and $d = 0.928$ for the 3- and 4-band condition, respectively). The remaining subjects (L222 and L282) performed at chance level across all vocoded conditions. Of the two subjects that had some degree of hearing impairment (decreased hearing at 4 kHz, see Methods), one subject (L307) exhibited above-average performance in the 3 and 4-band conditions, while the other (L282) performed at chance levels (see Fig. 1). Both of these subjects exhibited near-ceiling accuracy with clear speech.

Overall, across-subject median accuracy was significantly above chance for both 3- and 4-band conditions ($p = 0.00391$, $W = 44$, and $p = 0.00293$, $W = 53$, respectively, one-tailed Wilcoxon signed rank tests), with no systematic difference between the two ($p = 0.715$, $W = 19$, two-tailed Wilcoxon signed rank test). Accuracy in the 3-band condition was significantly greater than in the 2-band condition ($p = 0.00195$, $W = 0$), and accuracy in the clear condition was significantly greater than in the 4-band condition ($p = 0.00195$, $W = 0$; two-tailed Wilcoxon signed rank test).

For the purpose of relating the subjects' task performance to underlying auditory cortical activity, a cutoff of 67.5% average accuracy in 3- and 4-band conditions (i.e. across-subject median) was used to differentiate good from poor performance on the task (dotted line in legend of Fig. 1). There was no significant relationship between the subjects' hearing (as measured by better-ear pure tone average) and average accuracy in 3- and 4-band conditions ($p = -0.0982$, $p = 0.787$, Spearman's rank-order correlation).

RT decreased between the 2- and 3-band conditions ($p = 0.00977$, $W = 52$), and between the 4-band and clear stimuli ($p = 0.00195$, $W = 55$), but not between 1- and 2-band ($p = 0.625$, $W = 33$) or 3- and 4-band ($p = 1$, $W = 28$, two-tailed Wilcoxon signed rank tests). At a group level, decreases in RTs with less stimulus spectral degradation paralleled improvements in identification accuracy (Fig. 1, bottom panel). In summary, behavioral measures (accuracy, sensitivity and RTs) revealed the subjects' uniformly poor (floor) performance for 1 and 2-band condition, uniformly good (ceiling) performance with clear speech, and variable performance in the 3- and 4-band condition. There was no significant relationship between median reaction times and accuracy (3-band: $\rho = -0.334$, $p = 0.345$; 4-band: $\rho = -0.219$, $p = 0.544$; clear: $\rho = -0.442$, $p = 0.189$; Spearman rank correlation).

A potential confound when examining neural activity elicited by spectrally degraded speech is whether the physiology is strictly related to spectral complexity or intelligibility, as both properties co-vary with increase in the number of vocoder bands. If neural responses are based on spectral complexity, they are expected to monotonically increase from 1-band to 4-band to clear condition in both good and poor performers. In contrast, intelligibility is expected to be represented in good performers by a stepwise increase in activity from 1- and 2- band to 3- and 4-band condition, and from the latter to clear speech condition. In poor performers, intelligibility is expected to be represented by comparable neural responses to

all four levels of speech degradation, and a marked increase in responses elicited by clear speech.

3.2 Physiologic response patterns elicited by spectrally degraded speech

Auditory cortex exhibited regional differences in responses to spectrally degraded speech. To illustrate these regional differences, two subjects were chosen, L275 and B335, who demonstrated relatively good and poor performance, respectively. in the 3- and 4-band conditions. Figure 2 depicts responses from L275, who performed at chance in the 1- and 2-band conditions, had an average accuracy of 70% in the 3-and 4-band conditions, and reached near-ceiling identification accuracy with clear stimuli. Anatomical reconstructions are shown in Figure 2a. Sites labeled A through D traversed HG along its posteromedial to anterolateral axis. Sites E through H were located on progressively more anterior regions of the lateral STG.

Large responses, maximal in the high gamma frequency band, were elicited by all stimuli in the posteromedial portion of HG regardless of the degree of spectral degradation or intelligibility (Fig. 2b, sites A through C). In contrast, high gamma activity simultaneously recorded from anterolateral HG in this subject showed small increases as stimuli became less spectrally degraded and more intelligible, with the largest change occurring between 4-band and clear conditions (Fig. 2b, site D).

More varied response patterns were observed on the lateral STG (Fig. 2b, sites E through H). Activity in the posterior portion of lateral STG was time-locked to the onset of the sounds and did not exhibit a systematic relationship with stimulus condition (e.g. Fig. 2b, site E). In contrast, more anterior sites along STG (F and G) were characterized by low-magnitude responses elicited by the onsets of the unintelligible 1- and 2-band stimuli, and robust increases in high gamma activity between 2- and 3-band stimuli, and between 4-band and clear speech. These increases paralleled improvements in this subject's stimulus identification accuracy. Finally, the most anterior site H exhibited a marked increase in response magnitude to clear speech compared to any of the spectrally degraded stimuli, suggesting a posterior to anterior gradient along the lateral STG wherein responses become progressively more selective for clear speech.

The examples shown in Figure 2b were representative of the distribution of high gamma activity across the entire recording grid in this subject (Fig. 2c). Responses to 1- and 2-band stimuli were restricted to the most posterior sites and several loci in the more anterior portion of the gyrus. A progressive increase in activity over most of the STG was observed to the increasingly intelligible stimuli.

Examples of auditory cortical responses in a subject (B335) who exhibited poor performance on the task in the 3- and 4-band conditions are shown in Figure 3. This subject exhibited relatively poor task performance across all vocoded speech conditions ($d' < 1$) and responded with 100% identification accuracy for clear speech. Additionally, this subject, who was left hemisphere-dominant, clinically required bilateral electrode coverage that was comparable between the two hemispheres (Fig. 3). This coverage provided an opportunity to address the question of whether responses from the left hemisphere (as shown for subject

L275 in Fig. 2) are comparable with activity in the right hemisphere. In order to promote reasonable comparisons between the two hemispheres in subject B335, four pairs of electrodes were chosen based on similarity of their MNI coordinates and orientation to gross anatomical landmarks. Four sites in the dominant hemisphere (A through D) were thus compared with four sites in the non-dominant hemisphere (E through H) (Fig. 3a). Within posteromedial HG of both hemispheres (sites A and E), strong activation was elicited by all stimuli regardless of their spectral complexity or intelligibility (Fig. 3b). This pattern was similar to that observed in posteromedial Heschl's gyrus of the better-performing subject L275 (cf. Fig. 2b, sites A, B, C). While activity at more lateral HG sites in subject B335 (sites B and F) was larger than that observed in subject L275 (cf. Fig. 2. Site D), the same pattern of marked increase in activity to clear stimuli compared to spectrally degraded speech was observed in both subjects. Activity on the STG sites of both hemispheres (C, D, G, H) was selective for clear stimuli, in parallel with this subject's behavioral performance and contrasting with the more robust STG responses seen in the better-performing subject L275 (cf. Fig. 2b). In summary, there was no evidence to indicate hemispheric differences in processing of vocoded and clear speech stimuli.

3.3 High gamma responses to vocoded and clear stimuli along HG and STG

The exemplar data (see Figs. 2 and 3) illustrate the most salient response patterns elicited by clear and vocoded speech within HG and STG. To investigate whether high gamma responses elicited by clear and spectrally degraded speech systematically varied along the long axes of HG and STG, each gyrus was parcellated into three sections (medial-middle-lateral for HG, and posterior-middle-anterior for STG). Parcellating the data along the long axes of HG and STG was motivated by results from studies demonstrating a progressive response selectivity for intelligible speech within ever more anterior portions of temporal cortex (Scott et al., 2000, 2006; Okada et al., 2010; Evans et al., 2014). The three-way gyrus parcellations were also consistent with response gradients observed along both gyri in previous intracranial studies (Nourski et al., 2014, 2017). Thus, each recording site was assigned an ROI label based on its location along HG or STG (Fig. 4a).

Within HG, there was a progressive decrease in high gamma activity along the medial-to-lateral axis (Fig. 4b) across all stimulus conditions ($p < 0.001$, Wilcoxon rank sum tests on high gamma ERBP elicited by all stimuli between adjacent portions of HG). In our subject cohort, high gamma ERBP within the medial third of HG did not significantly change across conditions ($p > 0.05$, Wilcoxon signed rank tests on high gamma ERBP between adjacent stimulus conditions, e.g. 1- vs 2- and 2- vs. 3-band, FDR-corrected for multiple comparisons). Thus, a portion of presumed core auditory cortex responds strongly to the stimuli regardless of their spectral complexity or intelligibility.

Evidence for high gamma response selectivity to clear speech was found, however, in both the middle and lateral portions of HG. In these regions, there was a significant increase in high gamma ERBP elicited by clear stimuli compared to the 4-band condition. Curiously, there was a small but statistically significant ($p = 0.0466$) decrease in high gamma ERBP from 3- to 4-band condition in the middle third of HG. Overall, changes in high gamma ERBP in the middle and lateral third of HG qualitatively reflect changes in spectral

complexity seen between vocoded and clear stimuli. Further analysis of responses recorded from HG in subjects who had above- and below-average task performance was precluded by the relatively small sample sizes and considerable variability in HG recording site counts across the ten subjects (see Table 1). This limitation, however, was of less concern for the STG, where electrode coverage was present in each subject, and the overall sample sizes were more amenable to subdividing the data set based on the subjects' task performance.

On STG, overall power in the high gamma frequency range was comparable between its posterior and middle portions across the five stimulus conditions ($p > 0.05$), whereas activation was more modest in the anterior compared to the middle portion of STG (1-band: $p = 0.00185$; 2-band, 3-band, 4-band: $p < 0.001$; clear: $p = 0.0115$, Wilcoxon rank sum tests) (Fig. 4c). Significant increases in high gamma ERBP were reliably observed between 2- and 3-band conditions, and between 4-band and clear conditions, in all three STG ROIs. Additionally, a modest but significant ($p = 0.0131$; Wilcoxon signed rank test) increase in ERBP was observed between 3- and 4-band condition in the posterior portion of STG. This observation would be consistent with activity within this area primarily reflecting spectral complexity. Changes in high gamma responses across stimulus conditions seen in middle and anterior portions of STG, on the other hand, were only observed between unintelligible (2-band) and intelligible (3-band), and intelligible (4-band) and clear speech (cf. Fig. 1).

3.4 Relationship between high gamma activity and behavioral performance

The results shown in Figure 4 were based on data pooled across all subjects regardless of their task performance, and revealed no significant difference in cortical responses to the unintelligible 1- and 2- band stimuli in any of the ROIs. Increases in the stimulus spectral complexity were typically associated with significant changes in high gamma responses throughout the studied regions of auditory cortex. To examine the relationship between high gamma responses and task performance, single-trial high gamma ERBP was compared between trials associated with correctly and incorrectly identified 3-& 4-band-vocoded stimuli at each recording site. This analysis failed to reveal any recording sites where correctly identified stimuli were associated with significantly larger high gamma activity ($p > 0.05$ for all sites; two-sample t-tests between trials associated with correct and incorrect responses with FDR correction for multiple comparisons). This negative finding was observed even when analysis was restricted to subjects who exhibited above-average performance in the 3- and 4-band conditions.

Because no relationship between high gamma activity and task performance was found on a trial-bytrial and site-by-site basis, we next sought to examine relationships between across-trial average high gamma responses, cortical location, and the subjects' overall task performance as indexed by identification accuracy for the 3- & 4-band stimuli. These results are presented in Figure 5, which depicts average high gamma ERBP measured at each recording site separately as a function of the site's location along HG and STG- and for each subject (Fig. 5). Based on similar intelligibility as assessed by each subject's task performance (see Fig. 1), 1- & 2-band, and 3- & 4-band conditions were collapsed. As the largest across-subject variability in task performance was observed in the 3- & 4-band condition, the subjects were rank-ordered and color-coded based upon their average

performance in the 3- & 4-band condition (see legend of Fig. 5). Sites where high gamma power failed to significantly increase above baseline (one-sample one-tailed *t*-test, $p > 0.05$, FDR-corrected) were plotted on the bottom of each panel. No consistent relationship between high gamma ERBP and the subjects' task performance was observed within HG, where increases in the 3- & 4-band condition relative to the 1 & 2-band condition failed to cluster into groups based on subjects' task performance (Fig. 5a).

In contrast to HG, vocoded speech typically elicited little to no high gamma ERBP along STG in subjects who performed poorly on the task (depicted by cool colors in Fig. 5), while clear speech elicited comparable responses in all subjects regardless of their task performance. This latter finding is important as it (1) indicates that the sampled STG sites were capable of generating robust high gamma activity regardless of individual subjects' behavioral performance in the 3- & 4-band condition and (2) parallels the near-ceiling behavioral performance with clear speech in the subject cohort. Notably, in subjects who performed well on the task in the 3- & 4-band condition (warm colors in Fig. 5), activity in all three STG ROIs was robust even in the 1- & 2-band condition, where performance was at chance (Fig. 5b, left column). This finding indicates that lateral STG in better performing subjects is more activated by vocoded speech in general than poor performers.

The latter observation is emphasized in Figure 6, which compares high gamma activity across all electrode sites on the lateral STG as a function of good versus poor performance in the 3- & 4-band condition. The limited number of sites in HG across the 10 subjects precluded comparisons between good and poor performers. Neural responses to vocoded stimuli were significantly larger on the lateral STG in subjects who exhibited good task performance ($p < 0.001$, Wilcoxon rank sum test), while responses to clear speech were comparable between the two groups of subjects and were not significantly different ($p = 0.290$).

This general finding was examined in more detail by first comparing the two groups of subjects for each of the five stimulus conditions separately (Supplementary Fig. 2a). For each vocoded stimulus condition, activity on STG was larger in good performers compared to poor performers. Next, data from the two groups of subjects were compared separately for each ROI within lateral STG (Supplementary Fig. 2b). Again, statistical analysis showed that high gamma responses to vocoded stimuli were larger in good performers in each of the three ROIs on the lateral STG. Finally, STG electrodes were subdivided into two subgroups (dorsal and ventral) based upon the rotated coordinate z_θ that was orthogonal to the long axis of the STG (see Methods) (Supplementary Fig. 2c). This manipulation permits examination of responses from electrodes closer to the Sylvian fissure or closer to the superior temporal sulcus. This examination was motivated by the possibility that there was an organization based upon preferred noise bandwidth orthogonal to the long axis of the STG, as described in unit recordings in the rhesus monkey (Rauschecker & Tian, 2004). Comparisons between data contributed by good- and poor-performing subjects yielded the same statistical relationship as that when STG was subdivided along its length.

4 Discussion

4.1 Summary of findings

A major source of perceptual variability in CI users is thought to be based on differences in auditory processing at the cortical level (Giraud & Lee, 2007; Lazard et al., 2012; Finke et al., 2016). This study used noise-vocoded speech as a model of CI stimulation in normal-hearing listeners, while they performed identification of stop consonants based on place of articulation. Consistent with earlier reports on identification of spectrally degraded speech (Shannon et al., 1995; McGettigan et al., 2014), 1- and 2-band conditions resulted in chance-level accuracy, whereas clear speech yielded performance at or near ceiling. Crucially, 3- and 4-band-vocoded stimuli yielded variable performance across subjects. High gamma activity within medial portion of HG, a core auditory region, was similarly robust across all stimuli regardless of their spectral degradation. In contrast, a progressive preference for clear speech was found in the lateral portion of HG. Variability in subjects' task performance in the 3-and 4-band condition was related to the degree of activation on the lateral STG elicited by all vocoded stimuli while behavioral and neural responses to clear speech were comparable across all subjects.

4.2 Heschl's gyrus

Core auditory cortex within the posteromedial third of HG responds strongly to speech regardless of spectral degradation or subjects' behavioral performance. Absence of a direct relationship between high gamma ERBP and speech perception in this region has also been reported for temporally degraded (time-compressed) speech (Nourski et al., 2009). Similarly, fMRI studies have demonstrated strong activation within posteromedial HG to both clear and intelligible noise-vocoded speech as well as unintelligible spectrally rotated speech (Okada et al., 2010). The lack of correlation between activity within core auditory cortex with intelligibility extends to other types of degraded speech and appears to be a general rule of sound processing at this level (Davis & Johnsrude, 2003; Kyong et al., 2014). It remains to be determined whether single-unit or multi-unit analyses are more selective in their response patterns to intelligible speech or vocoded speech with less spectral degradation (Bitterman et al., 2008; Brosch et al., 2011; Niwa et al., 2012). The relatively sparse coverage of HG (compared to STG) precluded detailed statistical comparisons between good and poor performing subjects at this time and will require additional study.

Responses within the most anterolateral third of HG exhibited a marked enhancement to clear over vocoded speech. Population responses within this non-core region are generally weak to simple non-speech sounds such as click trains or regular-interval noise bursts (Brugge et al., 2009; Griffiths et al., 2010; Nourski et al., 2014). The underlying mechanisms responsible for these effects are unclear. However, lack of clear segregation of activity between good and poor performers in the 3- & 4- band condition suggests that responses here mainly reflect spectral complexity rather than stimulus intelligibility. What stimulus attributes are ultimately responsible for modulating activity in this region remains to be clarified.

Activity within the middle third of HG was intermediate in both magnitude and specificity for clear and vocoded speech. This observation suggests a gradual transition between core and belt processing stages along the gyrus, as opposed to a strict boundary between the two. This idea is supported by measurements of high gamma onset latencies to speech stimuli, which increase linearly along the gyrus (Nourski et al., 2014). Likewise, during induction of general anesthesia, phase-locked auditory responses to click trains persisted in the medial third of HG, whereas the middle third exhibited a progressive decrease in neural activity over the course of induction, and anterolateral HG was weakly responsive to these simple non-speech sounds even in the awake state (Nourski et al., 2017). Finally, dynamic causal modeling of effective connectivity within HG also supports a more nuanced three-way partition along its axis (Kumar et al., 2011).

Finally, tonotopic organization of HG (Moerel et al., 2014; Hackett, 2015; Leaver & Rauschecker, 2016) with excitatory and inhibitory response areas (Wu et al., 2008) can in part account for the differential high gamma responses to vocoded and clear speech along HG (see Fig. 5). As the spectral complexity of the stimuli increases, the potential contribution of inhibitory sidebands would be diminished, which in turn would lead to increased high gamma responses.

4.3 Superior temporal gyrus

Fundamental differences in response properties were observed on the lateral STG relative to those seen in the lateral portion of HG. As a rule, there was a more gradual increase in activity with spectral complexity (see Fig. 4). Of note, the majority of sites on the lateral STG that failed to exhibit increased in high gamma ERBP in the 3- & 4-band condition were contributed by subjects whose task performance was poor (see Fig. 5b). In contrast, no clear segregation based on performance was seen in the lateral HG (see Fig. 5a). While this dichotomy suggests a transformation of activity between HG and lateral STG (see Fig. 4c), it cannot be concluded that sites on the STG are responsive to intelligible speech, rather than merely reflecting sensitivity to spectral complexity (Leaver & Rauschecker, 2010). In contrast, most studies have implicated the STS as being preferentially responsive to intelligible speech (Scott et al., 2000, 2006; Davis and Johnsrude, 2003; Narain et al., 2003).

Responses to all vocoded speech stimuli were larger in better-performing subjects throughout the STG (see Fig. 6). Thus, high gamma activity on STG does not directly translate to correct vs. incorrect identification of vocoded speech stimuli, either on a trial by trial basis or in the average. In contrast, it reflects neural processing that is a prerequisite for correct identification, rather than a direct biomarker for it. In other words, strong high gamma activation on STG may or may not lead to comprehension; weak activation will not. This interpretation helps explain negative results of our single-trial-based analysis.

The difference in responses to vocoded speech on the STG was observed regardless of whether the STG was examined as a whole, divided into three ROIs along the length of the gyrus or into two groups based on their dorsal or ventral location within the gyrus (see Supplementary Fig. 2). Thus, a simple metric that examines the degree of activation of lateral STG by spectrally degraded speech can predict identification accuracy for our syllable identification task. This observation parallels findings obtained from multiple

neuroimaging studies in CI users, which show a correlation between the magnitude of speech-elicited activity within non-primary auditory cortex and speech recognition (Fujiki et al., 1999; Green et al., 2005; Mortensen et al., 2006). It is of note that in a case study of a patient with a long history of CI use and good speech recognition performance and who required invasive monitoring for her medically-intractable epilepsy, (Nourski et al., 2013b), patterns of activity in response to a range of auditory stimuli on the lateral STG were well-differentiated and comparable to those seen in normal-hearing listeners (Howard et al., 2000; Nourski et al., 2009, 2013a). In total, these observations support a working hypothesis that both the magnitude and the spatially distributed patterns of activity on lateral STG of CI users can predict the degree of accurate speech perception that the implant can afford.

It might reasonably be argued that differences in performance and brain activation were based on the degree of attention subjects allotted during the task. However, enhancement of high gamma ERBP across all vocoded conditions, but not clear speech, in good relative to poor performers mitigates against this simple explanation. Attention might be expected to exert a more global effect across all randomly ordered stimuli. If attentional effects were responsible for these differences, then poor performers would not have reached ceiling performance to clear speech, reaction times would have been longer in the poor performers, and high gamma activity would have been significantly less in the clear speech condition. Finally, experience with spectrally degraded stimuli, as encountered by CI users, would likely be associated with progressively enhanced activity on the lateral STG, as listeners become more perceptually adept with the degraded stimuli.

4.4 Caveats

One major caveat of the present study is related to the studied subject population, which is an issue of concern inherent to all human intracranial studies. Experimental subjects have a neurologic disorder, and their auditory cortical responses may not be representative of a healthy population (Nourski & Howard, 2015). To address this issue, recordings from epileptic foci in each subject were excluded from the analyses. Importantly, results were replicated in multiple subjects, who had different neurologic histories, seizure foci and antiepileptic drug regimens. Cognitive function in each subject was in the average range, and all subjects were able to perform the experimental task successfully. Finally, auditory cortical activity in each subject was examined under multiple other experimental protocols that failed to exhibit aberrant response patterns that, in turn, would have served as a warning to exclude those subjects from the cohort.

A second caveat is the relatively limited stimulus set used in this study. The two speech exemplars - /aba/ and /ada/ - provided a convenient experimental tool for establishing basic response patterns at the cortical level, yet represented only one of the many phonemic contrasts present in human speech. Further, a larger stimulus set with more spectrally degraded conditions would permit an intelligibility threshold to be established for each subject. This would allow for the intelligibility threshold to be used as a dependent variable in the analysis to further refine the complex relationship between spectral complexity and intelligibility. Additionally, a noise vocoder, as implemented in this study, is a relatively

coarse approximation of CI output to the central auditory system. Therefore, a more realistic acoustic model of CI output might lead to more nuanced results.

Finally, as can be appreciated from the time-frequency plots within Figures 2 and 3, ERBP changes in lower frequency bands were also observed. Changes within the low gamma frequency range (30-70 Hz) were similar to those seen in the high gamma band within both HG and STG (see Fig. 2, 3). The relationships between high gamma and the other frequency bands (beta: 14-30 Hz, alpha, 8-14 Hz; theta: 4-8 Hz) were more variable. There could be suppression of ERBP in the low frequency bands (e.g. theta) with or without concomitant increases in high gamma (sites C, D, G, H in Fig. 3b). Further, the timing of changes in these lower bands was typically more prolonged compared to that seen in high gamma. A detailed analysis of these relationships was beyond the scope of the paper and would require additional subjects to provide adequate statistical power for these multiple comparisons.

4.5   Concluding remarks

The current study did not observe hemispheric specialization within HG and lateral STG during the phonemic identification task using vocoder-degraded speech (see Fig. 3). Some previous non-invasive studies have demonstrated left-lateralized activity during processing of intelligible speech (Davis & Johnsrude, 2003; Narain et al., 2003; Peelle et al., 2013; Kyong et al., 2014), whereas others reported bilateral involvement of the temporal lobe (Millman et al., 2015; Lee et al., 2016; Wijayasiri et al., 2017). While the current study found comparable degrees of activation in non-core auditory cortex of both hemispheres, results should not imply that they each hemisphere is performing the same functions during speech processing (Meyer et al., 2002; Kyong et al., 2014; Lee et al., 2016). In more realistic scenarios of human-to-human communication, when spectral detail is degraded, semantic processing in the language-dominant hemisphere becomes engaged to assist in the deciphering of impoverished speech (Lee et al., 2016). The results of the present study cannot address this important hemispheric specialization due to the relatively simple phonemic task that did not require semantic processing (cf. McGettigan & Scott, 2012).

Within both HG and STG, ERBP was maximal in the high gamma frequency range. While ERBP in the low gamma frequency band was typically smaller than that seen in the high gamma band, the overall pattern of stepwise increases across stimulus conditions was similar (data available upon request). This similarity between these frequency bands may be translationally relevant, as the gamma band is more accessible with non-invasive EEG and MEG techniques. Thus, non-invasive studies looking at gamma activity may be able to infer activity patterns within the high gamma frequency band, which in turn has been used as a surrogate for multiunit activity (Ray et al., 2008; Steinschneider et al., 2008; Jenison et al., 2015). Relationships between increases in ERBP within the high gamma frequency band and concurrent decreases in the alpha and theta bands may also have translational relevance. Finally, the variability in the degree of auditory cortical activation in CI users may serve to customize rehabilitation training strategies used to improve CI performance.

The present study demonstrates that responses to vocoded speech are diminished in poor versus good performers, but clear speech is comparable between the two groups. The functional properties of auditory cortex on the lateral STG responsible for this effect are

unknown. Differences in intrinsic organizational features of non-primary auditory cortex, including sensitivity to amplitude/frequency modulation and stimulus bandwidth (Rauschecker & Tian, 2004; Kusmierek & Rauschecker, 2014), are all possible explanations for inter-subject performance variability. As these organizational properties are based on convergence of inputs from lower stages of the auditory hierarchy (e.g., Lee & Winer, 2011; Rauschecker, 2015), they would represent examples of bottom-up processing mechanisms.

Top-down influences on auditory cortical activity may also be important determinants (e.g. Davis et al., 2005; Hervais-Adelman et al., 2008; Bhargava et al., 2014; Clarke et al., 2016). As a rule, processing speech that is degraded by any number of causes (e.g. age- or noise-related hearing loss) engages non-auditory areas (Eisner et al., 2010; Obleser & Weisz, 2012; Wild et al., 2012). Listening to noise-vocoded speech recruits an extensive cortical network that includes superior temporal sulcus, middle temporal gyrus, prefrontal and premotor cortex (e.g. Davis & Johnsrude, 2003; Lee et al., 2016). The specific roles of these auditory-related areas in perception of spectrally degraded speech remain unknown. Future work aiming to define these roles will likely require tasks that engage lexico-semantic processing which may not be necessary for the simple phonemic stimuli used here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Abrams DA, Ryali S, Chen T, Balaban E, Levitin DJ, Menon V. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. Cereb Cortex. 2013 7;23(7):1703–14. [PubMed: 22693339]

Anderson CA, Lazard DS, Hartley DE. Plasticity in bilateral superior temporal cortex: Effects of deafness and cochlear implantation on auditory and visual speech processing. Hear Res. 2017 1;343:138–149 [PubMed: 27473501]

Benjamini Y, and Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57, 289–300.

Beyea JA, McMullen KP, Harris MS, Houston DM, Martin JM, Bolster VA, Adunka OF, Moberly AC. Cochlear Implants in Adults: Effects of Age and Duration of Deafness on Speech Recognition. Otol Neurotol. 2016 10;37(9):1238–45. [PubMed: 27466894]

Bhargava P, Gaudrain E, Ba kent D Top-down restoration of speech in cochlear-implant users. Hear Res. 2014 3;309:113–23. [PubMed: 24368138]

Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I. Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature. 2008 1 10;451(7175):197–201 [PubMed: 18185589]

Blamey P, Arndt P, Bergeron F, Bredberg G, Brimacombe J, Facer G, Larky J, LindstrÖm B, Nedzelski J, Peterson A, Shipp D, Staller S, Whitford L. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. Audiol Neurootol. 1996 Sep-Oct;1(5):293–306. [PubMed: 9390810]

Blamey P, Artieres F, Ba kent D, Bergeron F, Beynon A, Burke E, Dillier N, Dowell R, Fraysse B, Gallégo S, Govaerts PJ, Green K, Huber AM, Kleine-Punte A, Maat B, Marx M, Mawman D,

Mosnier I, O'Connor AF, O'Leary S, Rousset A, Schauwers K, Skarzynski H, Skarzynski PH, Sterkers O, Terranti A, Truy E, Van de Heyning P, Venail F, Vincent C, Lazard DS Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. Audiol Neurootol. 2013;18(1):36–47. [PubMed: 23095305]

Boersma P (2001). Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341–345.

Brosch M, Selezneva E, Scheich H. Representation of reward feedback in primate auditory cortex. Front Syst Neurosci. 2011 2 7;5:5. [PubMed: 21369350]

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA, 3rd. Coding of repetitive transients by auditory cortex on Heschl's gyrus. J Neurophysiol. 2009 10;102(4): 2358–74. [PubMed: 19675285]

Clarke J, Ba kent D, Gaudrain E Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech. J Acoust Soc Am. 2016 1;139(1):395–405. [PubMed: 26827034]

Da Costa S, van der Zwaag W, Marques JP, Frackowiak RS, Clarke S, Saenz M. Human primary auditory cortex follows the shape of Heschl's gyrus. J Neurosci. 2011 10 5;31(40):14067–75. [PubMed: 21976491]

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. J Neurosci. 2003 4 15;23(8):3423–31. [PubMed: 12716950]

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen. 2005 5;134(2):222–41. [PubMed: 15869347]

Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J Acoust Soc Am. 1997 10;102(4):2403–11. [PubMed: 9348698]

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J Neurosci. 2010 5 26;30(21):7179–86. [PubMed: 20505085]

Evans S, Kyong JS, Rosen S, Golestani N, Warren JE, McGettigan C, Mourão-Miranda J, Wise RJ, Scott SK. The pathways for intelligible speech: multivariate and univariate perspectives. Cereb Cortex. 2014 9;24(9):2350–61. [PubMed: 23585519]

Finke M, Büchner A, Ruigendijk E, Meyer M, Sandmann P. On the relationship between auditory cognition and speech intelligibility in cochlear implant users: An ERP study. Neuropsychologia. 2016 7 1;87:169–81 [PubMed: 27212057]

Fujiki N, Naito Y, Hirano S, Kojima H, Shiomi Y, Nishizawa S, Konishi J, Honjo I. Correlation between rCBF and speech perception in cochlear implant users. Auris Nasus Larynx. 1999 7;26(3):229–36. [PubMed: 10419029]

Giraud AL, Lee HJ. Predicting cochlear implant outcome from brain organisation in the deaf. Restor Neurol Neurosci. 2007;25(3-4):381–90. [PubMed: 17943013]

Green KM, Julyan PJ, Hastings DL, Ramsden RT. Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness. Hear Res. 2005 7;205(1-2):184–92. [PubMed: 15953527]

Griffiths TD, Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Brugge JF, Howard MA. Direct recordings of pitch responses from human auditory cortex. Curr Biol. 2010 6 22;20(12):1128–32 [PubMed: 20605456]

Hackett TA. Anatomic organization of the auditory cortex. Handb Clin Neurol. 2015;129:27–53. [PubMed: 25726261]

Harnsberger JD, Svirsky MA, Kaiser AR, Pisoni DB, Wright R, Meyer TA. Perceptual "vowel spaces" of cochlear implant users: implications for the study of auditory adaptation to spectral shift. J Acoust Soc Am. 2001 5;109(5 Pt 1):2135–45. [PubMed: 11386565]

Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP. Perceptual learning of noise vocoded words: effects of feedback and lexicality. J Exp Psychol Hum Percept Perform. 2008 4;34(2):460–74 [PubMed: 18377182]

Holden LK, Finley CC, Firszt JB, Holden TA, Brenner C, Potts LG, Gotter BD, Vanderhoof SS, Mispagel K, Heydebrand G, Skinner MW. Factors affecting open-set word recognition in adults with cochlear implants. Ear Hear. 2013 May-Jun;34(3):342–60. [PubMed: 23348845]

Howard MA, 3rd, Volkov IO, Granner MA, Damasio HM, Ollendieck MC, Bakken HE. A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. Technical note. J Neurosurg. 1996 1;84(1):129–32. [PubMed: 8613821]

Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF. Auditory cortex on the human posterior superior temporal gyrus. J Comp Neurol. 2000 1 3;416(1):79–92. [PubMed: 10578103]

Jenison RL, Reale RA, Armstrong AL, Oya H, Kawasaki H, Howard MA, 3rd. Sparse Spectro-Temporal Receptive Fields Based on Multi-Unit and High-Gamma Responses in Human Auditory Cortex. PLoS One. 2015 9 14;10(9):e0137915. [PubMed: 26367010]

Kovach CK, Gander PE. The demodulated band transform. J Neurosci Methods. 2016 3 1;261:135–54. [PubMed: 26711370]

Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Howard MA, 3rd, Friston KJ, Griffiths TD. Predictive coding and pitch processing in the auditory cortex. J Cogn Neurosci. 2011 10;23(10):3084–94. [PubMed: 21452943]

Kusmierek P, Rauschecker JP. Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. J Neurophysiol. 2014 4;111(8):1671–85. [PubMed: 24501260]

Kyong JS, Scott SK, Rosen S, Howe TB, Agnew ZK, McGettigan C. Exploring the roles of spectral detail and intonation contour in speech intelligibility: an FMRI study. J Cogn Neurosci. 2014 8;26(8):1748–63. [PubMed: 24568205]

Lazard DS, Vincent C, Venail F, Van de Heyning P, Truy E, Sterkers O, Skarzynski PH, Skarzynski H, Schauwers K, O'Leary S, Mawman D, Maat B, Kleine-Punte A, Huber AM, Green K, Govaerts PJ, Fraysse B, Dowell R, Dillier N, Burke E, Beynon A, Bergeron F, Ba kent D, Artières F, Blamey PJ Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. PLoS One. 2012;7(11):e48739. [PubMed: 23152797]

Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci. 2010 6 2;30(22):7604–12. [PubMed: 20519535]

Leaver AM, Rauschecker JP. Functional Topography of Human Auditory Cortex. J Neurosci. 2016 1 27;36(4):1416–28. [PubMed: 26818527]

Lee CC, Winer JA. Convergence of thalamic and cortical pathways in cat auditory cortex. Hear Res. 2011 4;274(1-2):85–94. [PubMed: 20576491]

Lee YS, Min NE, Wingfield A, Grossman M, Peelle JE. Acoustic richness modulates the neural networks supporting intelligible speech processing. Hear Res. 2016 3;333:108–117. [PubMed: 26723103]

Li T, Fu QJ. Perceptual adaptation to spectrally shifted vowels: training with nonlexical labels. J Assoc Res Otolaryngol. 2007 3;8(1):32–41. [PubMed: 17131213]

Macmillan NA, Creelman CD. Detection Theory: A User's Guide. Psychology Press, NY 2004, 512 p.

McGettigan C, Scott SK. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends Cogn Sci. 2012 5;16(5):269–76. [PubMed: 22521208]

McGettigan C, Rosen S, Scott SK. Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: implications for cochlear implantation. Front Syst Neurosci. 2014 2 25;8:18. [PubMed: 24616669]

Merzenich MM. Coding of sound in a cochlear prosthesis: some theoretical and practical considerations. Ann N Y Acad Sci. 1983;405:502–8. [PubMed: 6575672]

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum Brain Mapp. 2002 10;17(2):73–88. [PubMed: 12353242]

Millman RE, Johnson SR, Prendergast G. The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. J Cogn Neurosci. 2015 3;27(3):533–45. [PubMed: 25244119]

Moberly AC, Bates C, Harris MS, Pisoni DB. The Enigma of Poor Performance by Adults With Cochlear Implants. Otol Neurotol. 2016 12;37(10):1522–1528. [PubMed: 27631833]

Moerel M, De Martino F, Formisano E. An anatomical and functional topography of human auditory cortical areas. Front Neurosci. 2014 7 29;8:225. [PubMed: 25120426]

Mortensen MV, Mirz F, Gjedde A. Restored speech comprehension linked to activity in left inferior prefrontal and right temporal cortices in postlingual deafness. Neuroimage. 2006 6;31(2):842–52. [PubMed: 16459106]

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R. Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science. 2005 8 5;309(5736):951–4. [PubMed: 16081741]

Nagahama Y, Schmitt AJ, Nakagawa D, Vesole AS, Kamm J, Kovach CK, Hasan D, Granner M, Dlouhy BJ, Howard MA, 3rd, Kawasaki H. Intracranial EEG for seizure focus localization: evolving techniques, outcomes, complications, and utility of combining surface and depth electrodes. J Neurosurg. 2018 5 25:1–13.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex. 2003 12;13(12):1362–8. [PubMed: 14615301]

Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, Malach R. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr Biol. 2007 8 7;17(15):1275–85. [PubMed: 17686438]

Niwa M, Johnson JS, O'Connor KN, Sutter ML. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci. 2012 2 29;32(9):3193–210. [PubMed: 22378891]

Nourski KV, Howard MA, 3rd. Invasive recordings in the human auditory cortex. Handb Clin Neurol. 2015;129:225–44. [PubMed: 25726272]

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, 3rd, Brugge JF. Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci. 2009 12 9;29(49):15564–74. [PubMed: 20007480]

Nourski KV, Brugge JF, Reale RA, Kovach CK, Oya H, Kawasaki H, Jenison RL, Howard MA, 3rd. Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. J Neurophysiol. 2013 3;109(5):1283–95. [PubMed: 23236002]

Nourski KV, Etler CP, Brugge JF, Oya H, Kawasaki H, Reale RA, Abbas PJ, Brown CJ, Howard MA, 3rd. Direct recordings from the auditory cortex in a cochlear implant user. J Assoc Res Otolaryngol. 2013b 6;14(3):435–50. [PubMed: 23519390]

Nourski KV, Steinschneider M, McMurray B, Kovach CK, Oya H, Kawasaki H, Howard MA, 3rd. Functional organization of human auditory cortex: investigation of response latencies through direct recordings. Neuroimage. 2014 11 1;101:598–609. [PubMed: 25019680]

Nourski KV, Banks MI, Steinschneider M, Rhone AE, Kawasaki H, Mueller RN, Todd MM, Howard MA, 3rd. Electrocorticographic delineation of human auditory cortical fields based on effects of propofol anesthesia. Neuroimage. 2017 5 15;152:78–93. [PubMed: 28254512]

Obleser J, Weisz N. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex. 2012 11;22(11):2466–77. [PubMed: 22100354]

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex. 2010 10;20(10):2486–95. [PubMed: 20100898]

Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex. 2013 6;23(6):1378–87. [PubMed: 22610394]

Petersen B, Gjedde A, Wallentin M, Vuust P. Cortical plasticity after cochlear implantation. Neural Plast. 2013;2013:318521. [PubMed: 24377050]

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage. 2001 4;13(4):669–83. [PubMed: 11305896]

Rauschecker JP. Auditory and visual cortex of primates: a comparison of two sensory systems. Eur J Neurosci. 2015 3;41(5):579–85. [PubMed: 25728177]

Rauschecker JP, Tian B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol. 2004 6;91(6):2578–89. [PubMed: 15136602]

Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. J Neurosci. 2008 11 5;28(45):11526–36. [PubMed: 18987189]

Reddy CG, Dahdaleh NS, Albert G, Chen F, Hansen D, Nourski K, Kawasaki H, Oya H, Howard MA, 3rd. A method for placing Heschl gyrus depth electrodes. J Neurosurg. 2010 6;112(6):1301–7. [PubMed: 19663547]

Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000 12;123 Pt 12:2400–6. [PubMed: 11099443]

Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. J Acoust Soc Am. 2006 8;120(2): 1075–83. [PubMed: 16938993]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995 10 13;270(5234):303–4. [PubMed: 7569981]

Smalt CJ, Gonzalez-Castillo J, Talavage TM, Pisoni DB, Svirsky MA. Neural correlates of adaptation in freely-moving normal hearing subjects under cochlear implant acoustic simulations. Neuroimage. 2013 11 15;82:500–9. [PubMed: 23751864]

Steinschneider M, Fishman YI, Arezzo JC. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. Cereb Cortex. 2008 3;18(3):610–25. [PubMed: 17586604]

Sweet RA, Dorph-Petersen KA, Lewis DA. Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J Comp Neurol. 2005 10 24;491(3):270–89. [PubMed: 16134138]

Tyler RS, Preece JP, Lowder MW. The Iowa audiovisual speech perception laser video disc. Laser Videodisc and Laboratory Report, University of Iowa at Iowa City, Department of Otolaryngology – Head and Neck Surgery 1989.

Whittingstall K, Logothetis NK. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron. 2009 10 29;64(2):281–9 [PubMed: 19874794]

Wijayasiri P, Hartley DEH, Wiggins IM. Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. Hear Res. 2017 8;351:55–67. [PubMed: 28571617]

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS. Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci. 2012 10 3;32(40):14010–21. [PubMed: 23035108]

Wu GK, Arbuckle R, Liu BH, Tao HW, Zhang LI. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. Neuron. 2008 4 10;58(1):132–43. [PubMed: 18400169]

Highlights

- High gamma responses in posteromedial HG are similar across stimulus conditions

- A progressive preference for clear speech in anterolateral HG

- Responses on STG to vocoded speech are larger in better-performing subjects

- Behavioral and neural responses to clear speech are comparable across all subjects
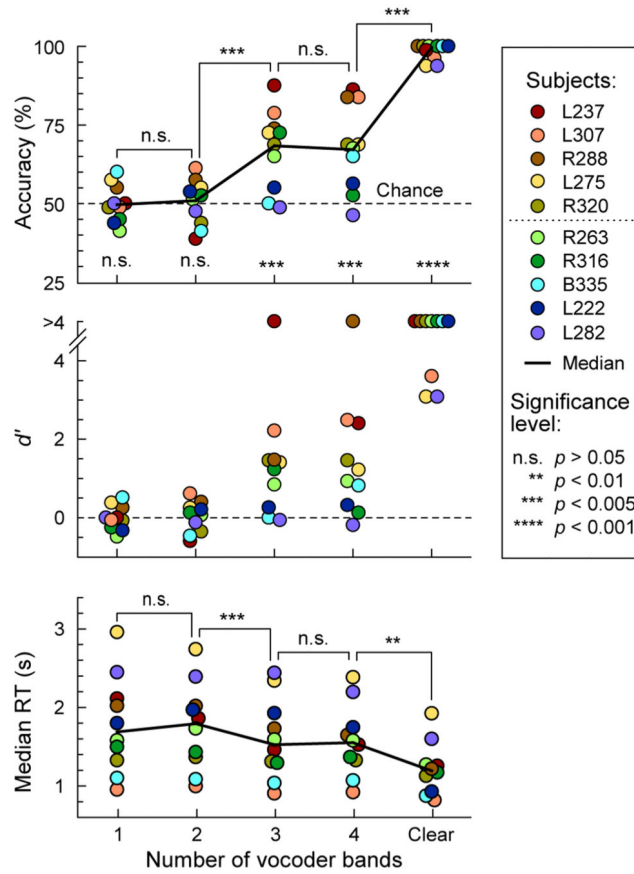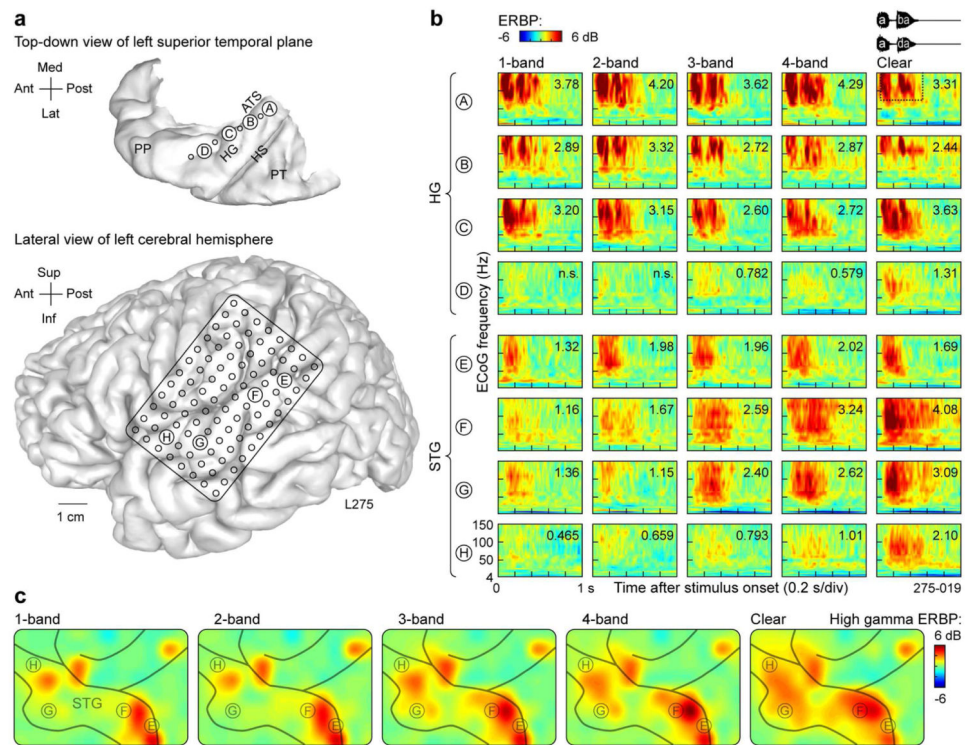
Figure 1.
Behavioral performance in the behavioral task. Task accuracy, sensitivity (*d′*) and RTs are plotted in the top, middle and bottom panels, respectively for vocoded (1-4 bands) and clear stimuli. Data from 10 subjects and across-subject median values are plotted as circles and solid lines, respectively. Significance of across-subject median task accuracy relative to chance performance (Wilcoxon signed-rank test) is indicated in the top panel. In the figure legend, subjects are rank-ordered based on their average accuracy in 3- and 4-band conditions, from best (L237) to worst (L282). Dotted line in figure legend corresponds to a cutoff of 67.5% accuracy, used to differentiate good from poor performance on the task.

Figure 2.
Responses to noise-vocoded and clear speech recorded from the auditory cortex in subject
L275, who exhibited good task performance. **a:** Top-down view of the superior temporal
plane and location of depth electrode contacts (left), and location of 96-contact subdural grid
implanted over perisylvian cortex (right). The superior temporal plane is aligned with the
lateral view of the left hemisphere along the anterior-posterior axis. ATS, anterior transverse
sulcus; HG, Heschl's gyrus; HS, Heschl's sulcus; PP, planum polare; PT, planum temporale.
b: Responses to noise-vocoded and clear sounds /aba/ and /ada/ (waveforms shown on top
right) recorded from sites A-D in HG (posteromedial to anterolateral) and sites E-H on
lateral STG (posterior to anterior). ERBP is computed for ECoG frequencies between 4 and
150 Hz. Numbers indicate high gamma ERBP values (dB relative to prestimulus baseline)
averaged within the time-frequency window for quantitative analysis (50-550 ms, 70-150
Hz; dotted rectangle in top right panel); "n.s." indicates instances where average high
gamma ERBP failed to reach significance (one-sample one-tailed t-test against zero, $p =$
0.05, FDR-corrected). **c:** Spatial distribution of high gamma activity simultaneously
acquired across the entire subdural electrode grid. A focal increase in high gamma activity
seen in postcentral gyrus (top right corner of the grid) likely represents somatosensory
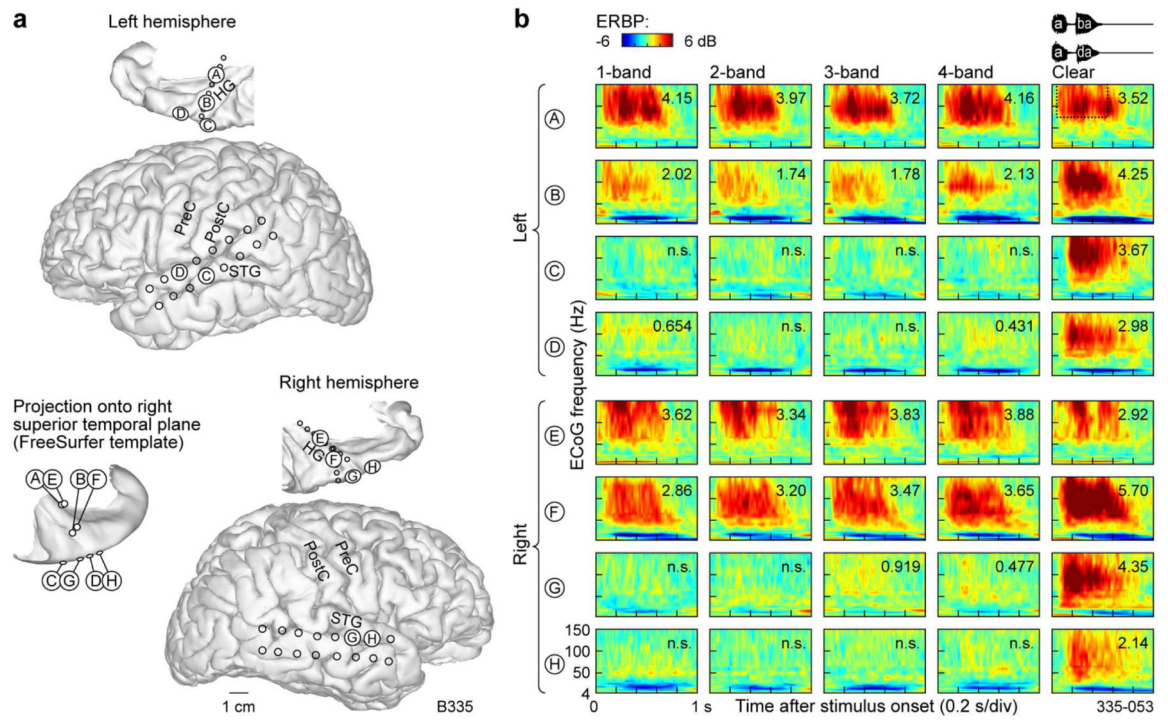feedback associated with the subject holding the game controller.

Figure 3.
Responses to noise-vocoded and clear speech recorded from the auditory cortex in subject B335, who exhibited poor task performance. See legend of Fig. 2 for details.
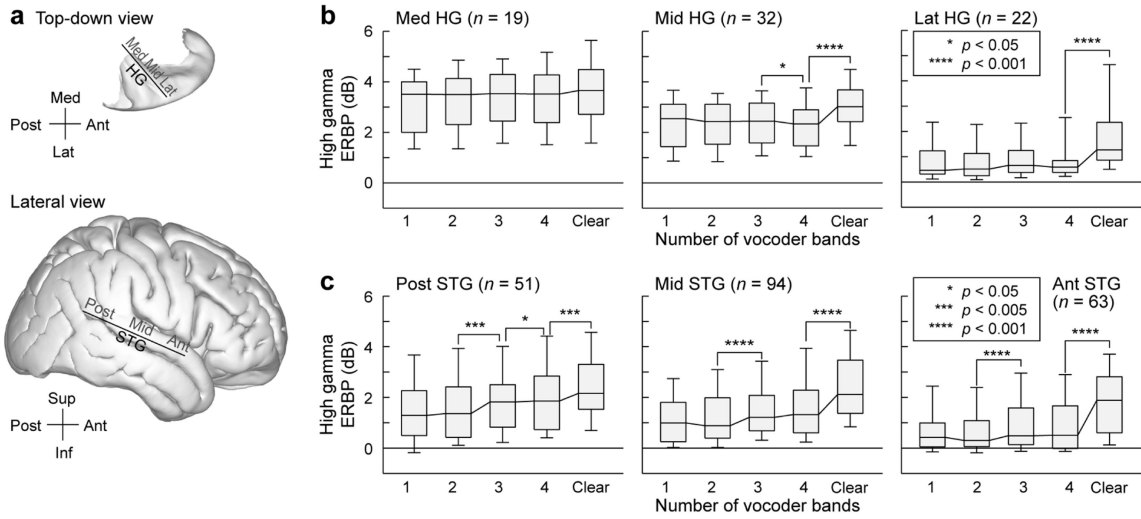
Figure 4.
High gamma ERBP elicited by vocoded and clear speech in HG and on STG. Data from 10 subjects. a: Top-down view of the right superior temporal plane and lateral view of right cerebral hemisphere (FreeSurfer average template brain). Solid lines denote the axes of coordinate rotation for HG and STG recording sites (x_θ and y_θ, respectively). b: High gamma ERBP recorded from medial, middle and lateral portions of HG (left, middle and right panels, respectively) in response to vocoded and clear stimuli. Box plots show across-electrode medians, quartiles, 10th and 90th percentiles. Stars represent significant differences between adjacent conditions, as determined by Wilcoxon signed rank tests with FDR correction for multiple comparisons. Numbers (*n*) on top of each panel represent the number of electrodes in each ROI. c: Same analysis, shown for electrodes within posterior, middle and anterior portions of STG (left, middle and right panel, respectively).
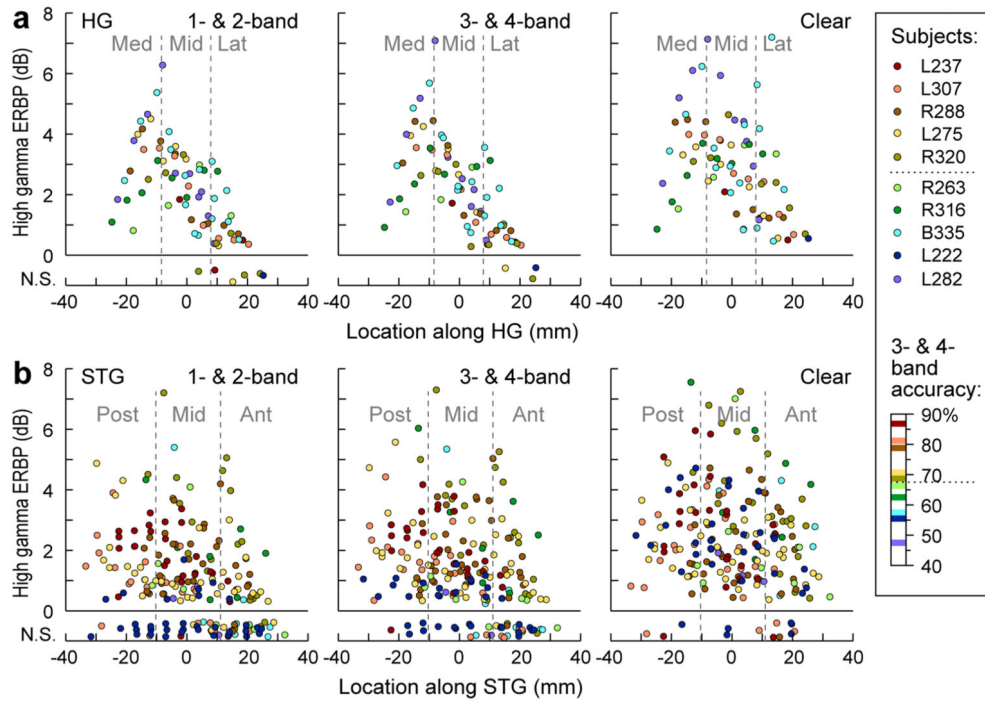
Figure 5.
High gamma ERBP across the length of HG and STG. Data from all electrodes in all 10 subjects. a: High gamma ERBP recorded from HG in response to 1- & 2-band vocoded, 3- & 4-band-vocoded and clear stimuli (left, middle and right panels, respectively) for each site, plotted as functions of location along HG. Sites where high gamma ERBP failed to reach significance (one-sample one-tailed *t*-test, $p = 0.05$, FDR-corrected) are plotted on the bottom of each panel, with jitter added to the Y-coordinate for clarity. Colors represent individual subjects, ranked by average identification accuracy in the 3- & 4-band condition (see Fig. 1). Dotted line in figure legend corresponds to a cutoff of 67.5% accuracy, used to differentiate good from poor performance on the task. b: Same analysis, shown for STG electrodes.
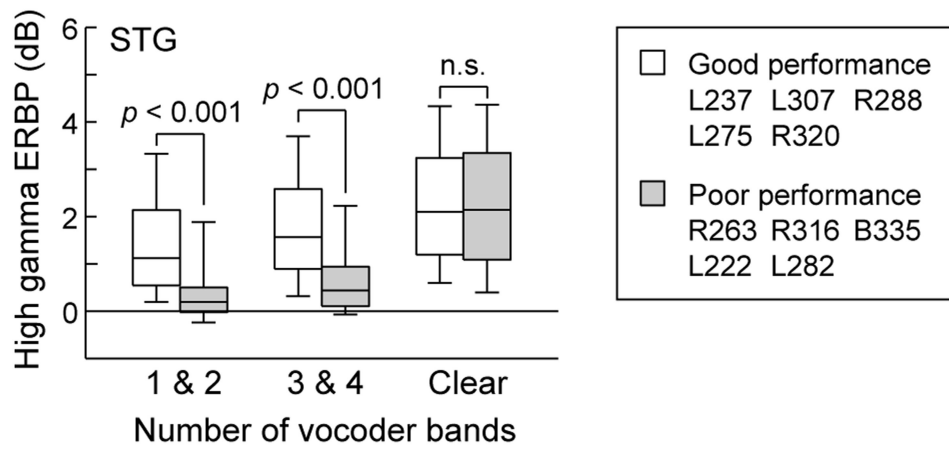
Figure 6.
Comparison of high gamma ERBP recorded from the STG in subjects who exhibited good vs. poor performance on the task, as measured by average accuracy in the 3- & 4-band condition (see text for details). Box plots show across-electrode medians, quartiles, 10th and 90th percentiles. P-values represent significant differences, as determined by Wilcoxon rank sum test.

Table 1.

Subject demographics.

Subject code	Age	Sex	No. of contacts		Seizure focus
			HG	STG	
L222	33	M	0	35	L medial temporal
L237	27	M	3	27	L anterior, medial and basal temporal
R263	32	F	4	8	R inferior lateral frontal, posterior ventral frontal, anterior insula
L275	30	M	8	38	L lateral ventral temporal cortex
L282	40	M	10	3	L posterior superior temporal, basal frontal
R288	20	M	8	32	R temporal pole, posterior suprasylvian cortex
L307	29	M	8	27	L posterior insula
R316	31	F	8	7	R medial temporal
R320	51	F	8	23	R medial temporal
B335	33	M	16	8	Bilateral medial temporal
Total			73	208	

F: female, HG: Heschl's gyrus; L: left; M: male, R: right; STG: superior temporal gyrus