

Predictive Top-Down Integration of Prior Knowledge during Speech Perception

Ediz Sohoglu, Jonathan E. Peelle, Robert P. Carlyon, and Matthew H. Davis

Medical Research Council Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom

A striking feature of human perception is that our subjective experience depends not only on sensory information from the environment but also on our prior knowledge or expectations. The precise mechanisms by which sensory information and prior knowledge are integrated remain unclear, with longstanding disagreement concerning whether integration is strictly feedforward or whether higher-level knowledge influences sensory processing through feedback connections. Here we used concurrent EEG and MEG recordings to determine how sensory information and prior knowledge are integrated in the brain during speech perception. We manipulated listeners' prior knowledge of speech content by presenting matching, mismatching, or neutral written text before a degraded (noise-vocoded) spoken word. When speech conformed to prior knowledge, subjective perceptual clarity was enhanced. This enhancement in clarity was associated with a spatiotemporal profile of brain activity uniquely consistent with a feedback process: activity in the inferior frontal gyrus was modulated by prior knowledge before activity in lower-level sensory regions of the superior temporal gyrus. In parallel, we parametrically varied the level of speech degradation, and therefore the amount of sensory detail, so that changes in neural responses attributable to sensory information and prior knowledge could be directly compared. Although sensory detail and prior knowledge both enhanced speech clarity, they had an opposite influence on the evoked response in the superior temporal gyrus. We argue that these data are best explained within the framework of predictive coding in which sensory activity is compared with top-down predictions and only unexplained activity propagated through the cortical hierarchy.

Introduction

It is widely acknowledged that our subjective experience reflects not only sensory information from the environment but also our prior knowledge or expectations (Remez et al., 1981; Rubin et al., 1997). A remarkable feature of the brain is its ability to integrate these two sources of information seamlessly in a dynamic and rapidly changing environment. However, the mechanisms by which this integration takes place are still unclear. One proposal is that perceptual processing is strictly feedforward, with sensory information and higher-level knowledge integrated at a postsensory decision stage in which multiple representations are evaluated before a final interpretation is selected (Fodor, 1983; Norris et al., 2000). An alternative account argues that sensory processing is directly modified by higher-level knowledge through feedback connections (McClelland and Elman, 1986; Friston, 2010).

Here we explore how sensory information and prior knowledge of speech content are integrated in the brain and modulate the subjective clarity of speech. Speech perception is an ideal context in which to study integration effects because in everyday listening we constantly exploit prior information—such as a speaker's lip movements or semantic context—to interpret incoming speech signals (Sumbly, 1954; Miller and Isard, 1963). Furthermore, the cortical network supporting speech perception is increasingly understood, showing a hierarchical organization that progresses from sensory processing in the superior temporal gyrus (STG) to more abstract linguistic and decision processes in the inferior frontal gyrus (IFG) (Scott and Johnsrude, 2003; Binder et al., 2004; Hickok and Poeppel, 2007). Given this anatomical organization, long-standing debates concerning whether speech perception is a purely feedforward process or includes feedback mechanisms can be construed in terms of functional interactions between the IFG and STG.

In the current study, listeners reported the subjective clarity of degraded spoken words. We manipulated prior knowledge of speech content by presenting matching, mismatching, or neutral text before speech onset. We also parametrically varied the level of speech degradation, and therefore the amount of speech sensory detail, so that changes in neural responses attributable to sensory information and prior knowledge could be directly compared. Because subjective experience of speech clarity is similarly enhanced by providing either more detailed sensory information or prior knowledge of speech content (Jacoby et al., 1988), we asked whether these two sources of enhanced subjective clarity have equivalent effects on neural responses.

Received Oct. 6, 2011; revised March 19, 2012; accepted April 20, 2012.

Author contributions: E.S., J.E.P., R.P.C., and M.H.D. designed research; E.S. performed research; E.S. analyzed data; E.S., J.E.P., R.P.C., and M.H.D. wrote the paper.

This research was supported by Medical Research Council Grant MC-A060-5PQ80. We are grateful to Maarten van Casteren, Clare Cook, and Lucy MacGregor for their assistance with data collection and also to Pierre Gagnepain, Olaf Hauk, Richard Henson, and Daniel Mitchell for their advice on MEG and EEG data analysis.

The authors declare no competing financial interests.

Correspondence should be addressed to either Ediz Sohoglu or Matthew H. Davis, Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK. E-mail: e.sohoglu@gmail.com, matt.davis@mrc-cbu.cam.ac.uk.

J. E. Peelle's present address Center for Cognitive Neuroscience and Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104.

DOI:10.1523/JNEUROSCI.5069-11.2012

Copyright © 2012 the authors 0270-6474/12/328443-11\$15.00/0

A critical test that can distinguish between bottom-up and top-down accounts is the timing of activity in sensory and higher-level regions (cf. Bar et al., 2006). We therefore combined high-density EEG and MEG recordings to obtain precise temporal and spatial measures of neural activity during speech perception. If a strictly bottom-up mechanism is involved in integrating sensory information and prior knowledge, sensory-related processing in the STG should be modulated by subjective clarity before abstract linguistic or decision computations in the IFG. Conversely, a top-down mechanism would be reflected by the opposite pattern, with abstract computations in the IFG being modulated before sensory-related processing in the STG.

Materials and Methods




Participants. Eighteen right-handed participants were tested after being informed of the procedure of the study, which was approved by the Cambridge Psychology Research Ethics Committee. All were native speakers of English, between 18 and 40 years old (mean \pm SD, 29 ± 6 years) and had no history of hearing impairment or neurological disease based on self-report. Data from four participants were excluded because of noisy EEG recordings (from high impedances or excessive eye/movement artifacts) resulting in 14 participants (eight female) in the final dataset.

Stimuli and procedure. A total of 324 monosyllabic words were presented in spoken or written format. The spoken words were 16 bit, 44.1 kHz recordings of a male speaker of southern British English, and their duration ranged from 317 to 902 ms (mean \pm SD, 598 ± 81 ms).

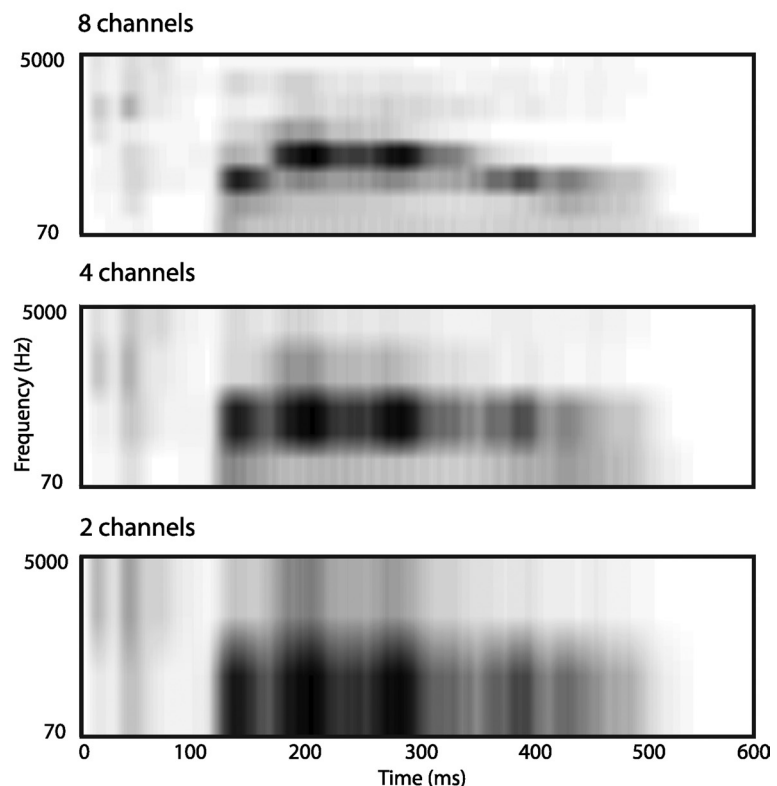
Prior knowledge of speech content was manipulated by presenting a written version of the spoken word before speech onset (matching condition) (Fig. 1A). Effects from matching written text were assessed relative to two control conditions in which prior knowledge was not informative with respect to upcoming speech. In the mismatching condition, the written word was different from the spoken word, and in the neutral condition, written text contained a string of “x” characters. Written words for the mismatching condition were obtained by permuting the word list for their spoken form. As a result, each written word in the mismatching condition was also presented as a spoken word and vice versa. Mean string length was equated across conditions. Written text was composed of black lowercase characters presented for 200 ms on a gray background.

The amount of sensory detail in speech was varied using a noise-vocoding procedure (Shannon et al., 1995), which superimposes the temporal envelope from separate frequency regions in the speech signal onto corresponding frequency regions of white noise. This allows parametric variation of spectral detail, with increasing numbers of channels associated with increasing perceptual clarity. Vocoding was performed using a custom MATLAB (MathWorks) script, using two, four, or eight spectral channels logarithmically spaced between 70 and 5000 Hz (Fig.

A Example word pairs

| | Written | Spoken |
|-------------|---------|---|
| Matching | clay |  |
| Mismatching | snail |  |
| Neutral | xxx |  |

B Spectrograms of example spoken word (clay)



C Trial events

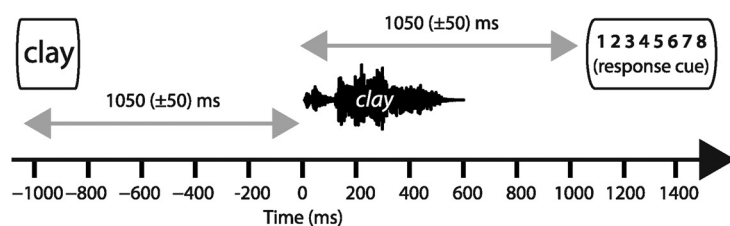


Figure 1. Stimulus characteristics. **A**, Example written–spoken word pairs used for matching, mismatching, and neutral conditions. **B**, Example spectrograms for the three speech sensory detail conditions. Speech with a greater number of spectral channels contained more sensory detail. **C**, Trial diagram showing the order and timing of events in each trial.

1B). Envelope signals in each channel were extracted using half-wave rectification and smoothing with a second-order low-pass filter with a cutoff frequency of 30 Hz. The overall RMS amplitude was adjusted to be the same across all audio files. Pilot data showed that mean \pm SD word report performance (across participants) at each of these sensory detail conditions is 3.41 ± 1.93 , 17.05 ± 1.98 , and $68.18 \pm 2.77\%$.

Manipulations of sensory detail (two-, four-, and eight-channel speech) and prior knowledge of speech content (matching/mismatching/neutral) were fully crossed, resulting in a 3×3 factorial design with 72

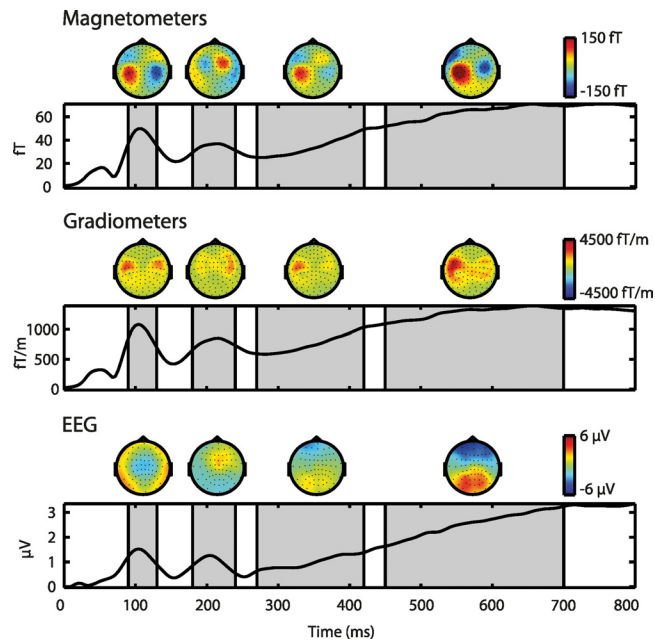


Figure 2. Time windows of the speech-evoked response over which the data were averaged before additional visualization and statistical analysis. Waveforms represent the global field power across sensors (after averaging across conditions and participants). Time windows are depicted by the areas shaded in gray. Topographic plots display the evoked response at each sensor averaged across time within each window.

trials in each condition. Trials were randomly ordered during each of four presentation blocks of 162 trials. For each participant, each of the spoken words appeared twice: either once as a matching trial and once as a mismatching trial, or twice as a neutral trial. The first presentation of each word occurred in the first two blocks of the experiment, and the second presentation occurred in the final two blocks. The particular words assigned to each condition were randomized over participants.

Stimulus delivery was controlled with E-Prime 2.0 software (Psychology Software Tools). Trials commenced with the presentation of a written word, followed 1050 ms later by the presentation of a spoken word (Fig. 1C). Participants were instructed to rate the clarity of each spoken word on a scale from 1 (“not clear”) to 8 (“very clear”). A response cue, which consisted of a visual display of the rating scale, was presented 1050 ms after the onset of the spoken word. Participants used a four-button box to navigate the rating scale and record their response. Subsequent trials began 850 ms after participants entered their responses. All time intervals were randomized by adding a random time of ± 0 –50 ms to reduce unwanted phase-locking of non-experimental factors (e.g., anticipatory responses). Before the experiment, participants completed a practice session of 18 trials containing all conditions but using a different corpus of words from those used in the main experiment.

Because of a software error, on $\sim 40\%$ of trials, the response cue displaying the rating scale was presented at the offset of the spoken word. Because the average duration of the spoken words was 598 ms, this meant that the response cue was presented on average 598 ms after speech onset, which is earlier than the intended timing of 1050 ms. We tested whether this erroneous timing of the response cue had any consequences for the speech-evoked neural response by including a factor of cue timing (“early response cue” or “late response cue”) for the 450–700 ms time window (for how time-windows were selected, see below, Sensor-space statistical analysis). This time window was tested because 97% of spoken words had durations > 450 ms. Therefore, for the majority of trials, the erroneous timing of the response cue could only have affected speech-evoked neural responses during this time window. A main effect of response cue timing was found over a small cluster of occipital MEG gradiometers ($p < 0.01$, FWE corrected), but this did not interact with the experimental manipulations of sensory detail and prior knowledge.

Data acquisition and pre-processing. Magnetic fields were recorded with a VectorView system (Elekta Neuromag) containing a magnetom-

eter and two orthogonal planar gradiometers at each of 102 positions within a hemispheric array. Electric potentials were simultaneously recorded using 70 Ag–AgCl sensors according to the extended 10–10% system and referenced to a sensor placed on the nose. All data were digitally sampled at 1 kHz and high-pass filtered above 0.01 Hz. Head position and EOG activity were continuously monitored using four head position indicator (HPI) coils and two bipolar electrodes, respectively. A 3D digitizer (Fastrak Polhemus) was used to record the positions of the EEG sensors, HPI coils, and ~ 70 additional points evenly distributed over the scalp, relative to three anatomical fiducial points (the nasion and left and right pre-auricular points).

Data from the MEG sensors (magnetometers and gradiometers) were processed using the temporal extension of Signal Source Separation (Taulu et al., 2005) in Maxfilter to suppress noise sources, compensate for motion, and reconstruct any bad sensors. Noisy EEG sensors were identified by visual inspection and excluded from additional analysis. Subsequent processing was done in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK) and FieldTrip (Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands) software implemented in MATLAB. The data were downsampled to 250 Hz and epoched -100 to 800 ms relative to speech onsets. Trials contaminated by EOG artifacts were removed by rejecting trials for which the amplitude in the 1–15 Hz range exceeded a set threshold of SD units from the mean across trials (established individually for each participant by visual inspection of the data). The remaining trials were low-pass filtered below 40 Hz and baseline corrected relative to the 100 ms pre-speech period, and the EEG data were referenced to the average over all EEG sensors. Epochs were averaged across trials to remove non-phase-locked activity and derive the evoked response.

Sensor-space statistical analysis. We restricted the search space for statistical analysis to portions of the evoked response when the signal-to-noise ratio (SNR) was high by averaging across time within each of four windows centered on prominent deflections in the evoked global field power (RMS amplitude over sensors). Those time windows are shown in Figure 2 and include the N100 (90–130 ms) and P200 (180–240 ms) components. For late latencies when there were no clear peaks, two broad windows were defined (270–420 and 450–700 ms) that correspond approximately to the early and late portions of the N400 component (cf. Desroches et al., 2009). After time averaging, F tests were performed across sensor space while controlling the FWE rate using random field theory (Kilner and Friston, 2010).

Before statistical analysis, the data were converted into 2D images by spherically projecting onto a 32×32 pixel plane for each epoch time sample and smoothed using a $5 \text{ mm} \times 5 \text{ mm} \times 10 \text{ ms}$ Gaussian kernel. In the case of gradiometers, an additional step involved combining the data across each sensor pair by taking the RMS of the two amplitudes. Results (condition means and error bars) are displayed by mapping statistically significant data points back onto the nearest corresponding sensor in the original head array.

Source reconstruction. To determine the underlying brain sources of the sensor data, a multimodal source inversion scheme was used to integrate data from all three neurophysiological measurement modalities (EEG and MEG magnetometers and gradiometers). This has been shown to give superior localization precision compared with considering each modality in isolation (Henson et al., 2009). To begin with, two separate forward models were constructed: one for the MEG sensors and another for the EEG. Both models had in common the use of a T1-weighted structural MRI scan obtained for each participant from which meshes

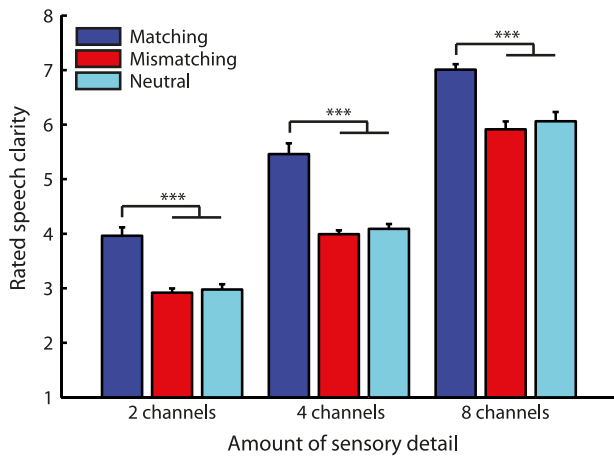


Figure 3. Behavioral results showing speech clarity ratings averaged across participants. The provision of increasing sensory detail and prior knowledge from matching text both led to an enhancement in perceived speech clarity. Error bars represent SEM across participants corrected for between-participant variability (Loftus and Masson, 1994). Braces show significance of *F* tests comparing matching with mismatching and neutral conditions (***) $p < 0.001$.

(containing 8196 vertices) were generated for the scalp and skull surfaces. Sensor locations and each participant's scalp mesh were then aligned using the digitized head shape. The MRI scan was also used to spatially transform a canonical cortical mesh in standard Montreal Neurological Institute (MNI) space to the individual space of each participant's MRI. To calculate the lead-field matrix, which specifies how any given source configuration will appear at the sensors, single-shell and boundary-element models were used for the MEG and EEG sensors, respectively. A parametric empirical Bayes framework (Phillips et al., 2005) was used for source inversion, using a LORETA (low-resolution brain electromagnetic tomography)-like approach (Pascual-Marqui, 2002), which attempts to minimize overall source power after initially assuming all elements are active and spatially correlated over adjacent regions. Multimodal fusion of the data was achieved by using a heuristic to convert all data to a common scale and by weighting each sensor type to maximize the model evidence (Henson et al., 2009). An additional constraint was imposed such that source solutions were consistent across participants, which has been shown to improve group-level statistical power (Litvak and Friston, 2008). Source power (equivalent to the sum of squared amplitude) in the 1–40 Hz range was derived from the resulting solutions and converted into 3D images.

Significant effects from sensor space were localized within the brain by averaging the 3D source power estimates across time within each window and mapping the data onto MNI space brain templates. Source estimates were subsequently converted into SNRs operationalized as statistical significance of pairwise *t* tests at the group level (i.e., mean signal divided by cross-participant variability). Given that the goal of source reconstruction was to localize the neural generators of sensor-space effects previously identified as significant, SNR maps are displayed with an uncorrected voxelwise threshold ($p < 0.05$).

Results

Behavioral results

Listeners' subjective ratings of speech clarity for each condition are shown in Figure 3. As expected, a repeated-measures ANOVA revealed that increasing sensory detail significantly enhanced speech clarity ($F_{(2,26)} = 298.62$, $p < 0.001$). Critically, prior knowledge of speech content provided by matching written text similarly enhanced spoken word clarity, relative to mismatching ($F_{(1,13)} = 91.72$, $p < 0.001$) or neutral ($F_{(1,13)} = 62.36$, $p < 0.001$) contexts. *Post hoc* comparisons revealed that this occurred even for two-channel speech, which contained the least amount of sensory detail (matching > mismatching, $t_{(13)} = 10.18$, $p <$

0.001; matching > neutral, $t_{(13)} = 6.71$, $p < 0.001$; Bonferroni's corrected for multiple comparisons). In addition, there was a small but significant decrease in clarity for mismatching compared with neutral contexts ($F_{(1,13)} = 5.13$, $p = 0.04$), indicating that incongruent prior knowledge can reduce speech clarity. However, the small magnitude of this effect suggests that incongruent prior knowledge has a lesser impact on subjective clarity than prior knowledge that is congruent with subsequent speech.

Sensor-space results

Sensors showing significant effects are shown in Figure 4, along with whole-head topographies expressing critical condition differences. Reported effects are all FWE rate corrected for multiple comparisons across sensors using a threshold of $p < 0.05$.

Significant main effects of speech sensory detail (Fig. 4A) were present 180–240 ms and later (270–420 and 450–700 ms) in the MEG (magnetometer and gradiometer) sensors but were absent in the EEG. The pattern of the means suggest that increasing sensory detail results in a larger evoked response.

To test for significant effects of prior knowledge from matching written text (Fig. 4B), a conjunction contrast (matching–mismatching AND matching–neutral) was used to detect significant differences in the evoked response between matching prior context and both mismatching and neutral contexts (Nichols et al., 2005). This conjunction contrast was motivated by our assumption that the effects of prior knowledge arise primarily when prior knowledge is congruent with speech, an assumption supported by our behavioral results showing that the main difference between our conditions lies between matching and the remaining mismatching/neutral conditions. Using a conjunction contrast also allowed us to control for expectation before speech onset (with the matching–mismatching contrast) while at the same time ruling out any minor effects from incongruent prior knowledge (with the matching–neutral contrast). Controlling for expectation before speech onset was a critical part of our design to ensure that we assessed only those effects of prior knowledge occurring after speech onset, because they reflect genuine integration of prior knowledge and incoming sensory information (cf. Arnal et al., 2009). Effects of prior knowledge from matching text were widespread in the EEG data, being present in all time windows, including the earliest 90–130 ms period. There were also effects in the magnetometers (270–420 and 450–700 ms) and gradiometers (450–700 ms). Although the EEG evoked response increased in the presence of matching written text, the opposite was true for the MEG sensors (i.e., the evoked response decreased). MEG effects of prior knowledge from matching text were also opposite in direction to the MEG sensory detail effects described previously (i.e., increased subjective clarity attributable to prior knowledge resulted in reduction rather than enhancement of the MEG response).

We additionally tested for effects of incongruent prior knowledge with the contrast mismatching–neutral. No significant effects were found in any sensor modality or time window, which further supports our decision to focus analysis on congruent effects of prior knowledge from matching text.

To further assess the relationship between neural and behavioral changes attributable to prior knowledge we conducted a single-trial analysis (Fig. 5) in which we correlated speech clarity ratings with the amplitude of the MEG and EEG signals in peak sensors showing an effect of prior knowledge (Fig. 5A; see also Fig. 4B). To avoid floor and ceiling effects in clarity ratings, we conducted this analysis only for responses to four-channel speech because this produced a range of clarity ratings for all participants

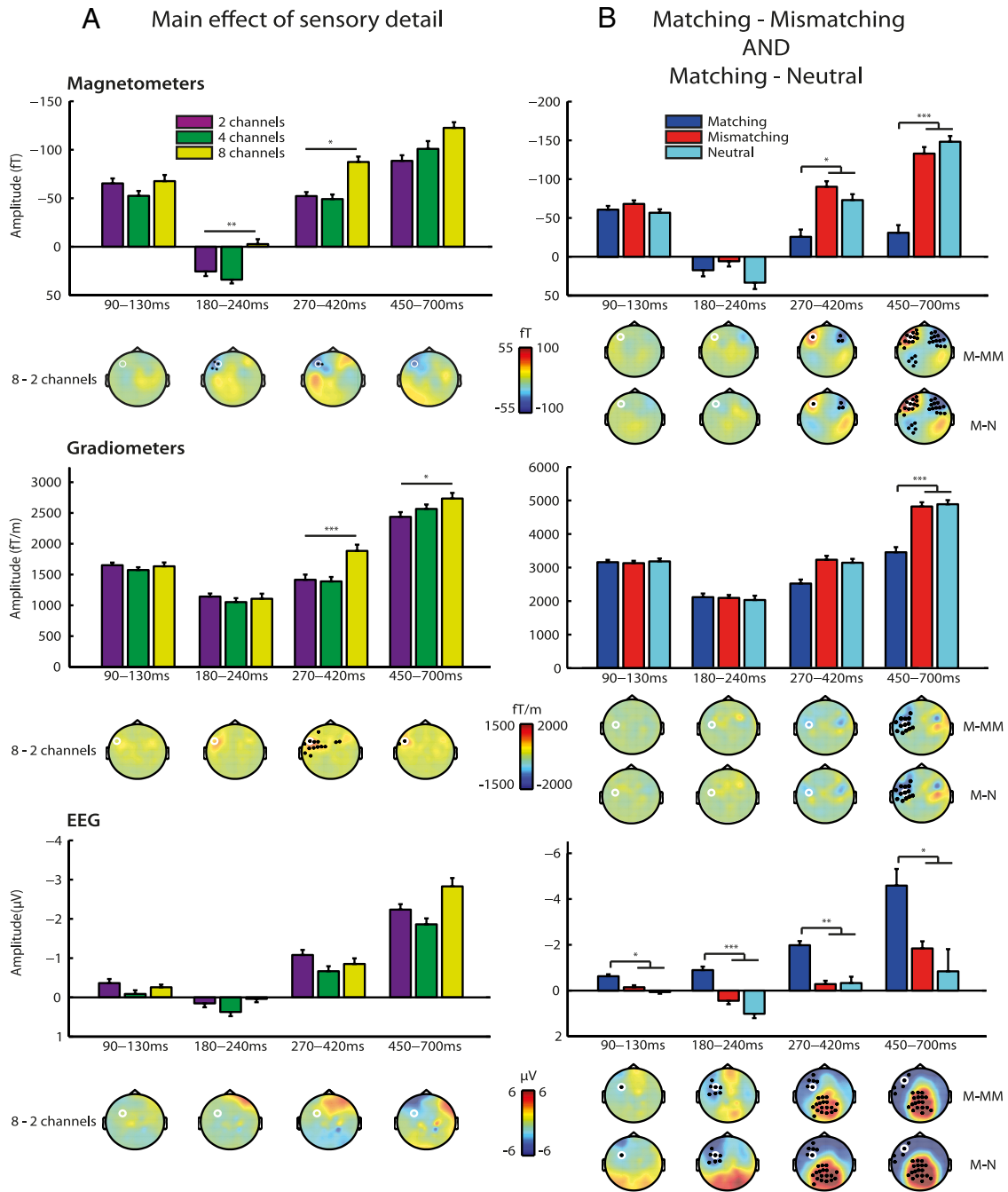


Figure 4. Speech-evoked response at selected sensors averaged across participants. **A**, Increasing speech sensory detail resulted in an enhancement of the evoked response. Error bars represent SEM across participants corrected for between-participant variability within each time window. Braces show time windows when there was a significant main effect of speech sensory detail for the sensor plotted ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$; FWE corrected for multiple comparisons across sensors). Topographic plots show the difference in response between eight-channel and two-channel speech. Small black dots on each topography depict locations of sensors showing significant effects for that time window, whereas large white circles depict locations of sensors from which signal has been plotted in the bar graph above. **B**, Prior knowledge from matching text resulted in a reduction of the evoked response for the magnetometer and gradiometer sensors and an enhancement for EEG sensors. Braces show time windows when there were significant differences between matching and both mismatching and neutral conditions (conjunction of matching–mismatching AND matching–neutral). Topographic plots show differences for matching–mismatching and matching–neutral conditions. M, Matching; MM, mismatching; N, neutral.

in all three prior knowledge conditions. For each participant and peak sensor, linear regression across single trials in each condition was used to quantify the change in neural response attributable to a single-unit change in rated speech clarity (cf. Lorch and Myers, 1990; Hauk et al., 2006) (for an example of data from one participant, see Fig. 5B). As shown in Figure 5C, slope estimates were significantly less than zero in the matching condition from 450 to 700 ms for a right frontal MEG magnetometer (two-tailed

paired t test, $t_{(13)} = -2.96$, $p < 0.05$; Bonferroni’s corrected across peak sensors). Regression slopes for mismatching and neutral conditions were not significantly different from zero after correcting for multiple comparisons but showed a positive trend in the neutral condition (two-tailed paired t test, $t_{(13)} = 2.50$, $p = 0.08$; Bonferroni’s corrected across peak sensors). In addition, there was a significant difference in regression slopes from 450 to 700 ms between the matching condition and the average of

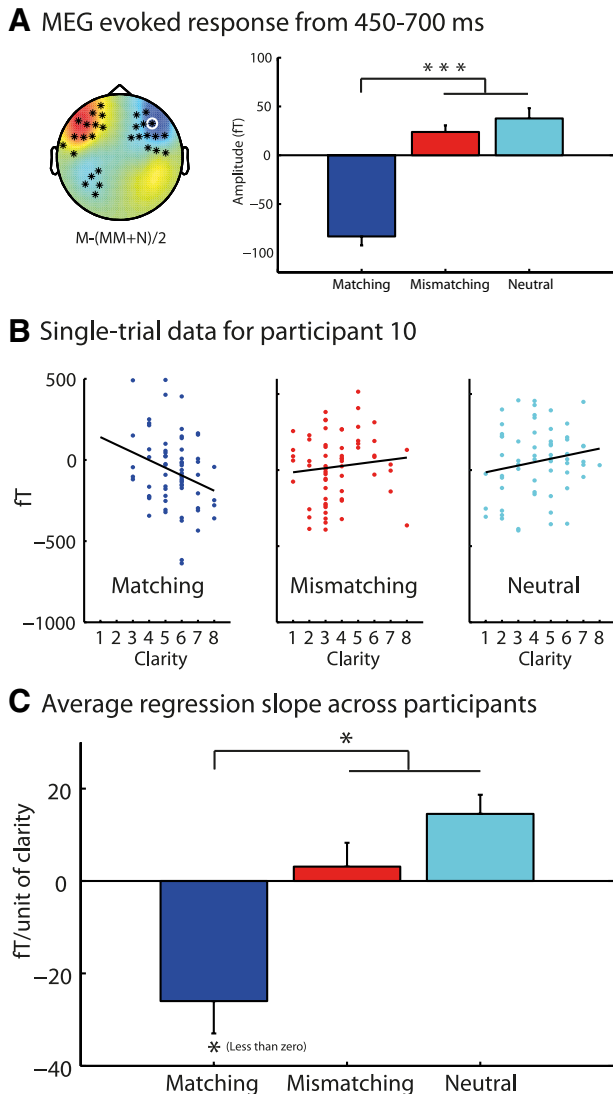


Figure 5. Single-trial analysis correlating behavioral ratings of subjective clarity and neural responses. **A**, Correlations were computed for sensors showing an effect of prior knowledge (matching–mismatching AND matching–neutral) on the averaged (evoked) response. The topographic plot shows the difference in evoked response between matching and average of mismatching and neutral trials from 450 to 700 ms. Small black dots indicate the locations of sensors showing significant effects of prior knowledge on the evoked response (as in Fig. 4B). The white circle indicates the location of the right frontal magnetometer in which significant correlations between behavioral and neural responses were found. The bar graph shows the average signal from this magnetometer for each prior knowledge condition averaged over sensory detail conditions. Error bars represent SEM across participants corrected for between-participant variability, and the brace shows significance of F test comparing matching with mismatching and neutral conditions ($***p < 0.001$; FWE corrected for multiple comparisons across sensors). **B**, Linear regression was used to compute the relationship between single-trial neural responses and clarity responses from each participant for the four-channel speech condition from 450 to 700 ms after speech onset. Graphs show data from a single participant. **C**, Regression slopes were significantly less than zero across participants for the matching condition and were significantly different between matching and the average of mismatching and neutral ($*p < 0.05$; Bonferroni’s corrected for multiple comparisons across sensors). M, Matching; MM, mismatching; N, Neutral.

mismatching and neutral conditions (two-tailed paired t test, $t_{(13)} = -3.32$, $p < 0.05$; Bonferroni’s corrected for multiple comparisons across peak sensors). This difference in correlation from prior knowledge parallels the pattern of the averaged (evoked) data with the regression slopes being reduced for the matching condition relative to mismatching and neutral conditions. This

further supports an association between reduced neural responses to degraded speech after matching written text and increases in subjective speech clarity.

Source localization of sensor-space effects

In the sensor-space analysis described above, EEG and MEG sensors showed different response profiles from prior knowledge: EEG showed an enhancement in signal amplitude, whereas MEG showed a reduction in amplitude. One explanation for this result is a differential sensitivity to cortical regions across EEG and MEG modalities. We therefore fused all three neurophysiological measurement modalities (EEG and MEG magnetometers and gradiometers) to obtain a single estimate of underlying cortical generators.

Figure 6A depicts the cortical generators of the effects of increasing speech sensory detail. At 180–240 ms, eight-channel speech showed greater source power than two-channel speech over the left STG and angular gyrus. During later time windows (270–420 and 450–700 ms), this activation extended onto the left middle and inferior temporal lobe.

As shown in Figure 6B, the earliest (90–130 ms) effect of prior knowledge from matching text reflected an increase in source power in a prefrontal cluster encompassing the left IFG and precentral and postcentral gyri, which shifted anteriorly for later time windows. From 270 to 420 ms, an additional increase in source power for the matching condition appeared over left middle occipital gyrus that was accompanied by additional increases from 450 to 700 ms over right middle temporal cortex and middle frontal gyrus (extending onto superior frontal gyrus; right-hemisphere effects are not shown in Fig. 6 but are listed in Table 1). The reduction of evoked response observed in MEG sensor space (that occurred from 270 to 420 ms and from 450 to 700 ms) only reached significance for the final time window from 450 to 700 ms. At this latency, there was a decrease in source power for the matching condition over the left STG.

Hence, our source reconstruction suggests that the EEG enhancement and MEG reduction effects of prior knowledge localize to the IFG and STG, respectively. This interpretation is consistent with previous studies showing that EEG and MEG provide complementary information on underlying cortical generators (Dale and Sereno, 1993; Sharon et al., 2007; Molins et al., 2008; Henson et al., 2009).

Our source reconstruction additionally suggests that prior knowledge from matching text modulates activity in a higher-order cortical region (left IFG) before peri-auditory cortex (left STG). This is precisely the finding predicted by a top-down mechanism. Furthermore, the effect of prior knowledge on STG activity is opposite in direction to the effect of sensory detail. We tested for this pattern of results directly by conducting a separate analysis in which we defined cortical regions of interest (ROIs) based on the source power averaged across all time windows. We selected a left frontal ROI by searching for voxels in which activity was greater for the matching condition relative to the average of mismatching and neutral contexts, using a voxelwise threshold of $p < 0.001$ (uncorrected). This revealed a cluster centered on the left IFG [peak at $(-42, 28, 26)$, $Z = 3.36$], extending into the left middle frontal gyrus (Fig. 7). A similar search was conducted to define a left temporal ROI, for an effect in the opposite direction (reduced activity for matching condition), which revealed a cluster centered on the left STG [peak at $(-56, -22, 4)$, $Z = 3.70$].

The graphs in Figure 7 depict differences in source power between matching and the average of mismatching and neutral conditions for the ROIs defined above. We entered these data

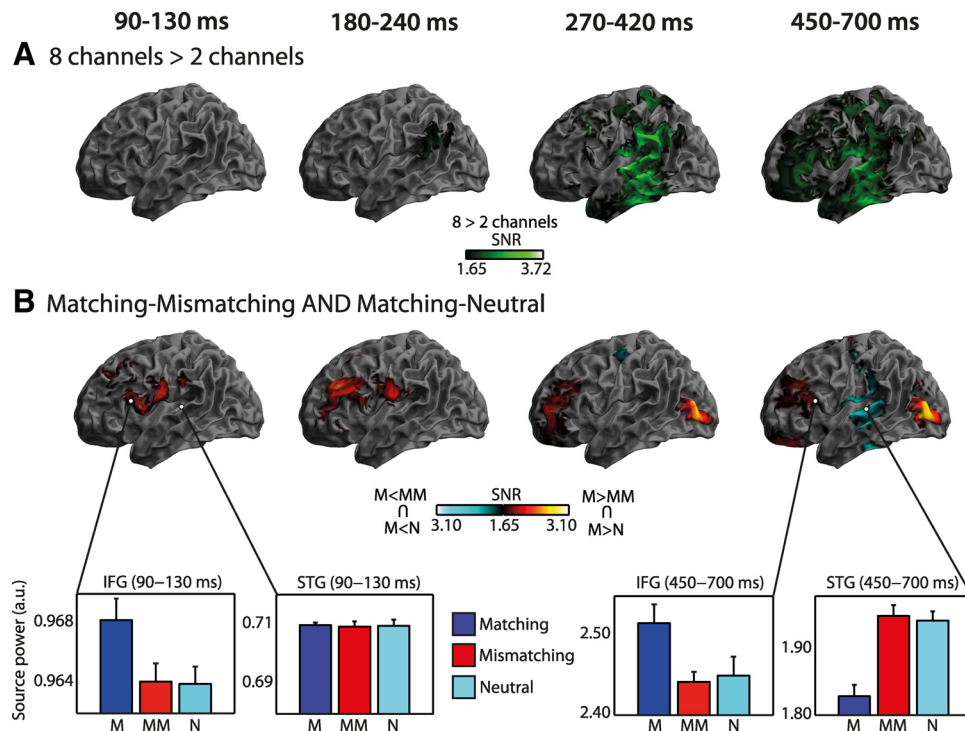


Figure 6. Source power SNRs (in units of Z-scores) for critical contrasts overlaid onto MNI space template brains (only left hemisphere shown). **A**, Brain regions showing greater source power for increasing sensory detail (8-channel > 2-channel speech). **B**, Brain regions showing differences in source power attributable to prior knowledge from matching text (conjunction of matching-mismatching AND matching-neutral). Bar graphs show source power (before noise normalization) for a voxel in the left IFG ($-54, 18, 20$) and left STG ($-56, -24, 4$). Error bars represent SEM across participants corrected for between-participant variability. M, Matching; MM, mismatching; N, neutral.

into a repeated-measures ANOVA with the factors time window (90–130 ms/450–700 ms) and region (IFG/STG). This revealed the expected main effect of region ($F_{(1,13)} = 40.18, p < 0.001$), as well as a significant interaction between time window and region ($F_{(1,13)} = 32.03, p < 0.001$). This confirms that activity in the IFG was modulated before activity in the STG, consistent with a top-down process.

Figure 7 additionally confirms that sensory detail and prior knowledge of speech content have differential effects on the evoked response. In the IFG (top graph), increasing sensory detail and the presence of prior knowledge similarly enhanced the evoked response. In the STG (bottom graph), the effect of prior knowledge had the opposite effect on evoked responses compared with the effect of sensory detail such that the evoked response was reduced. This pattern was tested directly by incorporating the effect of sensory detail (8–2 channels) into the ANOVA described above so that the origin of changes to the evoked response (sensory detail/prior knowledge) could be specified as a factor. This revealed a significant three-way interaction between origin of clarity, time window, and region ($F_{(1,13)} = 28.71, p < 0.001$).

Discussion

Our brains constantly integrate incoming sensory information and prior knowledge to produce a unified perceptual experience. In the current study, we investigated the way in which these different sources of information are rapidly combined during perception of degraded speech. We manipulated both sensory detail and prior knowledge of speech content and show that they similarly enhance speech clarity, in accordance with previous behavioral studies (Jacoby et al., 1988). Critically, by exploiting the hierarchical organization of the cortical speech network into sen-

sory and more abstract linguistic processing (Scott and Johnsrude, 2003; Hickok and Poeppel, 2007; Peelle et al., 2010), we demonstrate that the spatiotemporal profile of neural responses when prior knowledge facilitates speech perception is uniquely consistent with top-down modulation: effects of matching written text on speech processing occur in the IFG, a region associated with more abstract processing of speech content, before they occur in lower-level sensory cortex (STG).

The IFG has been implicated in amodal phonological analysis for written and spoken input alike (Price, 2000; Booth et al., 2002; Burton et al., 2005; Hickok and Poeppel, 2007) and shows increased responses when speech is degraded compared with when it is clear (Davis and Johnsrude, 2003; Shahin et al., 2009). Anatomically, data from nonhuman primates (Hackett et al., 1999; Romanski et al., 1999; Petrides and Pandya, 2009) and convergent evidence from functional connectivity (Anwander et al., 2007; Frey et al., 2008; Saur et al., 2008) show reciprocal connections between auditory and prefrontal cortices (including the IFG). These findings make the IFG well suited as the source of a top-down process whereby prior knowledge of abstract phonological content in speech (derived from prior matching text) interacts with lower-level acoustic-phonetic representations in the STG.

The early differential engagement of the IFG at 90–130 ms for degraded speech that follows matching text suggests that the effects of prior knowledge on auditory processing occur subsequent to an initial feedforward sweep that rapidly projects a representation of incoming speech to higher-level stages in the processing hierarchy. Given the distorted nature of the speech presented and the paucity of phonetic cues in the first 100 ms of a spoken word (only the initial consonant and vowel will have been

Table 1. Peak voxel locations (in MNI space) and summary statistics from source analysis

| Contrast | Time window | Region | Voxels (<i>n</i>) | Coordinates (mm) | Z |
|-------------------------|--------------|------------------------------|---------------------|------------------------------|-------|
| 8 > 2 channels | 180–240 ms | Left angular gyrus | 643 | −54, −60, 26 | 2.53 |
| | | Superior temporal gyrus | | −60, −46, 14 | 2.41 |
| | | Angular gyrus | | −40, −62, 26 | 2.09 |
| 8 > 2 channels | 270–420 ms | Left superior temporal gyrus | 6184 | −60, −20, 2 | 5.12 |
| | | Superior temporal gyrus | | −52, −22, 12 | 4.72 |
| | | Middle temporal gyrus | | −64, −36, 0 | 4.67 |
| | | Left postcentral gyrus | 1420 | −24, −36, 62 | 3.63 |
| | | Postcentral gyrus | | −22, −32, 72 | 2.92 |
| | | Middle frontal gyrus | | −44, 14, 42 | 2.58 |
| | | Right middle temporal pole | 1493 | 40, 16, −34 | 3.31 |
| | | Inferior temporal gyrus | | 42, −2, −44 | 2.99 |
| | | Superior temporal pole | | 34, 4, −18 | 2.91 |
| | | Right supramarginal gyrus | 1008 | 38, −38, 44 | 2.69 |
| | | Inferior parietal lobule | | 42, −52, 42 | 2.39 |
| | | Inferior parietal lobule | | 38, −44, 38 | 2.37 |
| | | 8 > 2 channels | 450–700 ms | Left superior temporal gyrus | 11763 |
| Superior temporal gyrus | −52, −20, 12 | | | 4.38 | |
| Superior temporal gyrus | −46, −20, 6 | | | 4.02 | |
| Left calcarine gyrus | 540 | | | −8, −94, −6 | 2.66 |
| Lingual gyrus | | | | −20, −94, −20 | 2.58 |
| Calcarine gyrus | | | | −6, −92, −14 | 2.30 |
| M > MM AND M > N | 90–130 ms | Left postcentral gyrus | 740 | −62, −10, 20 | 2.74 |
| | | Rolandic operculum | | −54, 4, 6 | 2.56 |
| | | Inferior frontal gyrus | | −54, 18, 20 | 2.30 |
| M > MM AND M > N | 180–240 ms | Left inferior frontal gyrus | 1151 | −42, 30, 26 | 3.37 |
| | | Middle frontal gyrus | | −38, 34, 32 | 3.28 |
| | | Inferior frontal gyrus | | −48, 22, 30 | 3.06 |
| M > MM AND M > N | 270–420 ms | Left middle occipital gyrus | 651 | −42, −84, 8 | 3.90 |
| | | Left inferior frontal gyrus | | −36, 38, 12 | 2.63 |
| | | Middle frontal gyrus | 1451 | −34, 36, 26 | 2.56 |
| | | Inferior frontal gyrus | | −36, 28, 20 | 2.28 |
| M > MM AND M > N | 450–700 ms | Left middle occipital gyrus | 906 | −42, −84, 10 | 4.06 |
| | | Middle occipital gyrus | | −48, −78, 0 | 3.76 |
| | | Middle temporal gyrus | | −48, −68, 16 | 2.23 |
| | | Right middle temporal gyrus | 3091 | 58, −40, −14 | 3.58 |
| | | Middle temporal gyrus | | 60, −54, 6 | 3.56 |
| | | Middle temporal gyrus | | 62, −42, 0 | 3.48 |
| | | Right middle frontal gyrus | 1468 | 32, 4, 52 | 3.15 |
| | | Superior frontal gyrus | | 28, 0, 46 | 3.00 |
| | | Middle frontal gyrus | | 30, 14, 50 | 2.99 |
| | | Left middle frontal gyrus | 1130 | −36, 36, 30 | 2.41 |
| | | Inferior frontal gyrus | | −50, 24, 18 | 2.38 |
| | | Inferior frontal gyrus | | −36, 28, 20 | 2.34 |
| | | M < MM AND M < N | 450–700 ms | Left superior temporal gyrus | 1637 |
| Postcentral gyrus | −58, −14, 16 | | | 3.64 | |
| Rolandic operculum | −50, −26, 14 | | | 3.60 | |

For display purposes, activations have been thresholded voxelwise at $p < 0.05$ (uncorrected) and clusterwise at $k > 500$ voxels (p values are uncorrected because statistical tests were performed in sensor space, corrected for multiple comparisons across sensors). M, Matching; MM, mismatching; N, neutral.

heard; cf. Grosjean, 1980), a coarse-grained representation of speech is likely to be involved. Nonetheless, there is sufficient information present in the speech signal to detect correspondence between current speech input and prior expectations. We propose that an early frontal mechanism detects the correspondence between written and spoken inputs rapidly after speech onset before the emergence of top-down processing. In the current study, this emergent top-down processing is reflected by concurrent modulation of inferior frontal and superior temporal cortices in later time windows from 270 to 420 ms and from 450 to 700 ms. This interpretation is consistent with a previous study using audiovisual speech showing that early activation in the IFG predicts the ensuing percept (Skipper et al., 2007), as well as studies using static auditory and visual stimuli showing that feedback processing manifests during later portions (>200 ms) of the

evoked response (Bar et al., 2006; Garrido et al., 2007; Epshtein et al., 2008). Strikingly, however, whereas top-down feedback can easily change ongoing processing for static stimuli because the sensory input is stationary and unchanging over the critical time period, our findings demonstrate ongoing top-down processes that track a dynamic and rapidly changing auditory signal. Abstract phonological information (from prior matching text) is in our work shown to modulate ongoing auditory processing of degraded speech.

One striking finding from our study is that sensory detail and matching prior knowledge of speech had comparable effects on subjective experience (increasing the perceived clarity of degraded speech) but had opposite effects on the evoked response in the STG. Whereas increasing sensory detail led to an enhancement in the evoked response (i.e., a larger response for eight-

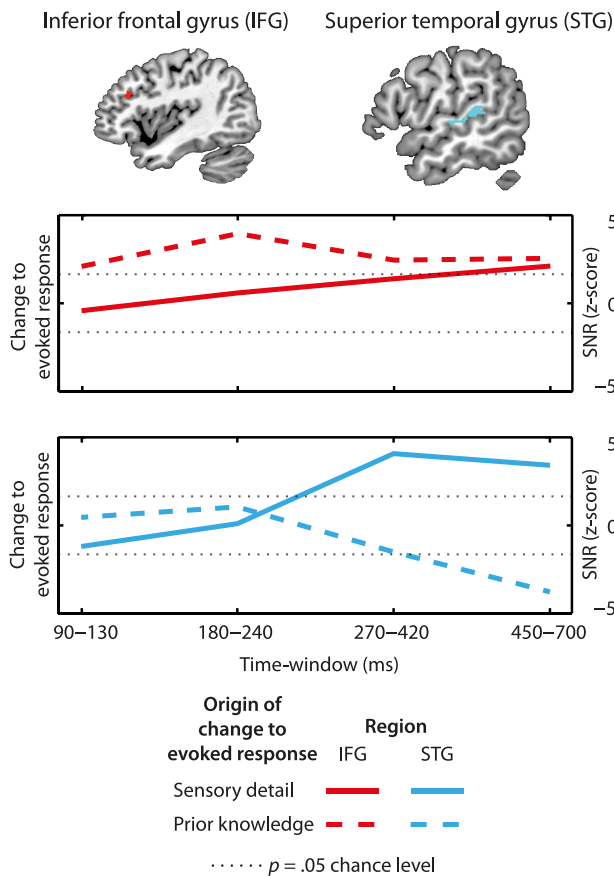


Figure 7. Results from ROI analysis showing changes to the evoked response that originate from sensory detail (8–2 channel) and prior knowledge of speech content (matching–average of mismatching and neutral). In the IFG [peak at (–42, 28, 26)], increasing sensory detail and prior knowledge from matching text similarly enhanced the evoked response (top graph). In the STG [peak at (–56, –22, 4)], the effect of prior knowledge had the opposite effect on the evoked response compared with the effect of sensory detail such that the evoked response was reduced (bottom graph). The two horizontal dotted lines in black denote a significance threshold of $p = 0.05$ between which changes in source power from sensory detail and prior knowledge were not significantly different from zero.

channel than for two-channel vocoded speech), the provision of prior knowledge reduced activity in the STG. The increased response for speech with more sensory detail is consistent with a number of previous studies that have shown increased hemodynamic (Davis and Johnsrude, 2003; Scott et al., 2006) and neurophysiological (Luo and Poeppel, 2007; Obleser and Kotz, 2011) responses for more spectrally detailed vocoded speech. However, the few studies that have shown changes in neural activity attributable to prior knowledge have typically observed increased responses (Hannemann et al., 2007) that may arise from prefrontal cortex (Giraud et al., 2004; Hervais-Adelman et al., 2012) or both prefrontal and auditory areas (Wild et al., 2012). Thus, our finding of opposite effects of sensory detail and prior knowledge in the STG is without precedent in previous studies of the perception of degraded speech and is inconsistent with accounts in which any enhancement to the perceived clarity of speech is accompanied by a corresponding increase in STG activity. It is, however, in agreement with the finding that recall of degraded spoken words from echoic memory is determined solely by the fidelity of sensory input rather than perceived clarity from top-down influences (Frankish, 2008). This suggests that, although sensory information and prior knowledge both enhance perceptual clarity, their effects can be dissociated by other behavioral

measures and, as demonstrated here, by neural responses in the STG.

One prominent model of speech perception that includes feedback connections is TRACE (McClelland and Elman, 1986; McClelland et al., 2006), which proposes hierarchically organized layers of localist units that represent speech using increasingly abstract linguistic representation (acoustic–phonetic features, phonemes, and words). A distinctive feature of this model is the presence of bidirectional connections between adjacent layers that allow prior lexical or phonological knowledge to influence ongoing phonological or acoustic–phonetic processes. This architecture would at least superficially make this model well suited to explaining the phonological–auditory interaction we are proposing. However, in the TRACE model, sensory and top-down inputs converge onto a single set of representational units (e.g., acoustic–phonetic units would be activated by both sensory and top-down phonological input). Hence, assuming that greater activation of model units corresponds to greater neural activity, TRACE would predict equivalent neural responses to changes in perceptual clarity caused by either sensory or top-down manipulations. Because we saw opposite effects of sensory detail and prior knowledge manipulations in the STG, we suggest that this form of feedback is challenged by the present results.

A second class of computational model that appears better able to account for the opposite effect of sensory and prior knowledge manipulations seen in our results is a form of hierarchical Bayesian inference termed “predictive coding.” This account, which is gathering increasing experimental support (Murray et al., 2002; van Wassenhove et al., 2005; Alink et al., 2010; Arnal et al., 2011), proposes that top-down predictions are compared with incoming sensory input and only unexplained activity (or error) propagated through the remainder of the processing hierarchy (Rao and Ballard, 1999; Friston, 2010). In the current context, we propose that abstract phonological predictions in the IFG (that originate from prior written text) are conveyed to the STG as acoustic–phonetic predictions that are then compared with neural representations of incoming speech input. Within this framework, listening conditions in which top-down predictions can explain a larger portion of sensory activity (such as when speech follows matching text) would result in less error and a reduction in activity, as seen in the STG in the present study. Conversely, speech with more sensory detail (i.e., eight-channel vs two-channel speech) should result in increased neural responses, because more spectrotemporal information is present in the signal that needs to be processed. Thus, we argue that our results are best described by predictive coding accounts, which propose comparison of top-down predictions with sensory input rather than the simple addition of top-down and sensory activation proposed in TRACE.

Although this predictive coding account of how prior knowledge modulates speech clarity is compelling, we acknowledge that the observed effects of matching text may be hard to distinguish from other aspects of listeners’ perceptual processing that change concurrently with speech clarity, such as their level of attention. Indeed, the precise relationship between predictive coding and attention is the subject of ongoing debate (cf. Summerfield and Egnor, 2009). Although this possibility cannot be ruled out completely, we observed that MEG responses to speech after matching text were significantly correlated with trial-by-trial variation in rated clarity and that this differed from the relationship seen for trials without matching prior knowledge. Furthermore, effects of prior knowledge and sensory detail occurred in the same brain regions (IFG and STG) and with a similar time course (e.g.,

in the STG, both effects co-occur from 270 to 700 ms). These observations suggest that the effect of prior knowledge is to modulate the same neural processes that respond to changes in sensory detail and that generate listeners' perceptual experience of speech clarity. Future work using transcranial magnetic stimulation or other invasive methods will be required, however, to show that there is a causal relationship between modulation of activity in the IFG and STG and changes to speech clarity.

In conclusion, our data provide evidence that prior knowledge is integrated with incoming speech through top-down feedback from the IFG to the STG. They additionally suggest that the neural impact of prior knowledge is opposite to the effect of more detailed sensory input despite both manipulations having the same impact on subjective perceptual clarity. These results suggest that the subjective experience of speech is governed by a general principle of predictive coding in which top-down predictions are compared with incoming sensory input and only unexplained activity propagated through the cortical hierarchy.

References

- Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30:2960–2966.
- Anwander A, Tittgemeyer M, von Cramon DY, Friederici AD, Knösche TR (2007) Connectivity-based parcellation of Broca's area. *Cereb Cortex* 17:816–825.
- Arnal LH, Morillon B, Kell CA, Giraud AL (2009) Dual neural routing of visual facilitation in speech processing. *J Neurosci* 29:13445–13453.
- Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14:797–801.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A* 103:449–454.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295–301.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002) Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage* 16:7–22.
- Burton MW, Locasto PC, Krebs-Noble D, Gullapalli RP (2005) A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage* 26:647–661.
- Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J Cogn Neurosci* 5:162–176.
- Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Desroches AS, Newman RL, Joanisse MF (2009) Investigating the time course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity. *J Cogn Neurosci* 21:1893–1906.
- Epshtein B, Lifshitz I, Ullman S (2008) Image interpretation by a single bottom-up top-down cycle. *Proc Natl Acad Sci U S A* 105:14298–14303.
- Fodor J (1983) *The modularity of mind*. Cambridge, MA: Massachusetts Institute of Technology.
- Frankish C (2008) Precategorical acoustic storage and the perception of speech. *J Mem Lang* 58:815–836.
- Frey S, Campbell JS, Pike GB, Petrides M (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci* 28:11435–11444.
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138.
- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A* 104:20961–20966.
- Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004) Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex* 14:247–255.
- Grosjean F (1980) Spoken word recognition processes and the gating paradigm. *Percept Psychophys* 28:267–283.
- Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45–58.
- Hannemann R, Obleser J, Eulitz C (2007) Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Res* 1153:134–143.
- Hauk O, Davis MH, Ford M, Pulvermüller F, Marslen-Wilson WD (2006) The time course of visual word-recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30:1383–1400.
- Henson RN, Mouchlianitis E, Friston KJ (2009) MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage* 47:581–589.
- Hervais-Adelman A, Carlyon RP, Johnsrude IS, Davis MH (2012) Motor regions are recruited for effortful comprehension of noise-vocoded words: evidence from fMRI. *Lang Cogn Process*. In press.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Jacoby LL, Allan LG, Collins JC, Larwill LK (1988) Memory influences subjective experience: noise judgments. *J Exp Psychol* 14:240–247.
- Kilner JM, Friston KJ (2010) Topological inference for EEG and MEG. *Ann Appl Stat* 4:1272–1290.
- Litvak V, Friston K (2008) Electromagnetic source reconstruction for group studies. *Neuroimage* 42:1490–1498.
- Loftus GR, Masson MEJ (1994) Using confidence intervals in within-subject designs. *Psychonom Bull Rev* 1:476–490.
- Lorch RF Jr., Myers JL (1990) Regression analyses of repeated measures data in cognitive research. *J Exp Psychol* 16:149–157.
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn Psychol* 18:1–86.
- McClelland JL, Mirman D, Holt LL (2006) Are there interactive processes in speech perception? *Trends Cogn Sci* 10:363–369.
- Miller G, Isard S (1963) Some perceptual consequences of linguistic rules. *J Verbal Learn Verbal Behav* 2:217–228.
- Molins A, Stufflebeam SM, Brown EN, Hämäläinen MS (2008) Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation. *Neuroimage* 42:1069–1077.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99:15164–15169.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Norris D, McQueen JM, Cutler A (2000) Merging information in speech recognition: feedback is never necessary. *Behav Brain Sci* 23:299–325; discussion 325–370.
- Obleser J, Kotz SA (2011) Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage* 55:713–723.
- Pascual-Marqui RD (2002) Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol* 24 [Suppl D]:5–12.
- Peelle JE, Johnsrude IS, Davis MH (2010) Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci* 4:51.
- Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol* 7:e1000170.
- Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ (2005) An empirical Bayesian solution to the source reconstruction problem in EEG. *Neuroimage* 24:997–1011.
- Price CJ (2000) The anatomy of language: contributions from functional neuroimaging. *J Anat* 197:335–359.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Remez RE, Rubín PE, Pisoni DB, Carrell TD (1981) Speech perception without traditional speech cues. *Science* 212:947–949.
- Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403:141–157.
- Rubin N, Nakayama K, Shapley R (1997) Abrupt learning and retinal size specificity in illusory-contour perception. *Curr Biol* 7:461–467.
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, Umarova

- R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A* 105:18035–18040.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Scott SK, Rosen S, Lang H, Wise RJ (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J Acoust Soc Am* 120:1075–1083.
- Shahin AJ, Bishop CW, Miller LM (2009) Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44:1133–1143.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Sharon D, Hämäläinen MS, Tootell RB, Halgren E, Belliveau JW (2007) The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. *Neuroimage* 36:1225–1235.
- Skipper JI, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387–2399.
- Sumby WH (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
- Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13:403–409.
- Taulu S, Simola J, Kajola M (2005) Applications of the signal space separation method. *IEEE Trans Signal Process* 53:3359–3372.
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181–1186.
- Wild CJ, Davis MH, Johnsrude IS (2012) Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage* 60:1490–1502.