Perceptual learning of degraded speech by minimizing prediction error

Ediz Sohoglu[a,1,2] and Matthew H. Davis[a,2]

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

Edited by Barbara Anne Dosher, University of California, Irvine, CA, and approved January 29, 2016 (received for review November 26, 2015)

Human perception is shaped by past experience on multiple timescales. Sudden and dramatic changes in perception occur when prior knowledge or expectations match stimulus content. These immediate effects contrast with the longer-term, more gradual improvements that are characteristic of perceptual learning. Despite extensive investigation of these two experience-dependent phenomena, there is considerable debate about whether they result from common or dissociable neural mechanisms. Here we test single- and dual-mechanism accounts of experience-dependent changes in perception using concurrent magnetoencephalographic and EEG recordings of neural responses evoked by degraded speech. When speech clarity was enhanced by prior knowledge obtained from matching text, we observed reduced neural activity in a peri-auditory region of the superior temporal gyrus (STG). Critically, longer-term improvements in the accuracy of speech recognition following perceptual learning resulted in reduced activity in a nearly identical STG region. Moreover, short-term neural changes caused by prior knowledge and longer-term neural changes arising from perceptual learning were correlated across subjects with the magnitude of learning-induced changes in recognition accuracy. These experience-dependent effects on neural processing could be dissociated from the neural effect of hearing physically clearer speech, which similarly enhanced perception but increased rather than decreased STG responses. Hence, the observed neural effects of prior knowledge and perceptual learning cannot be attributed to epiphenomenal changes in listening effort that accompany enhanced perception. Instead, our results support a predictive coding account of speech perception; computational simulations show how a single mechanism, minimization of prediction error, can drive immediate perceptual effects of prior knowledge and longer-term perceptual learning of degraded speech.

perceptual learning | predictive coding | speech perception | magnetoencephalography | vocoded speech

Successful perception in a dynamic and noisy environment critically depends on the brain's capacity to change how sensory input is processed based on past experience. Consider the way in which perception is enhanced by accurate prior knowledge or expectations. Sudden and dramatic changes in subjective experience can occur when a distorted and otherwise unrecognizable perceptual object is seen or heard after the object's identity is revealed (1–4). Such effects occur almost immediately; striking changes in perceptual outcomes occur over a timescale of seconds or less. However, not all effects of past experience emerge as rapidly as these effects of prior knowledge. With perceptual learning, practice in perceiving certain types of stimuli results in gradual and incremental improvements in perception that develop over a timescale of minutes or longer (Fig. 1*A*) (5, 6, 7). Critically, perceptual learning can generalize beyond the stimuli experienced during training, e.g., to visual forms presented in different retinal positions or orientations or spoken words that have not been heard before (6, 8–10). Thus, perceptual learning may have great potential in ameliorating sensory deficits (11–13), and understanding the neural and computational mechanisms supporting learning is critical.

Although prior knowledge and perceptual learning are both experience-dependent forms of perceptual improvement, the distinct time courses of their effects suggest that they originate in different brain mechanisms. Dual-mechanism accounts therefore propose that the influence of prior knowledge resides at a hierarchically late (e.g., decision) stage of processing and attribute the effect of learning to offline synaptic changes in earlier-level sensory cortex that take place after sensory stimulation (14, 15). However, other work has shown that perceptual learning of degraded stimuli is enhanced if accurate prior knowledge is provided before the presentation of degraded or otherwise ambiguous sensory input (6, 14, 16–18). Consistent with this behavioral association, alternative single-mechanism accounts have proposed that a single system containing multiple interacting levels of representation supports the effects of both prior knowledge and perceptual learning (15, 19, 20). According to these single-mechanism accounts, abstract higher-level representations derived from prior knowledge are used to inform and guide earlier, lower-level sensory processes. These interactions not only modulate immediate perceptual outcomes but also lead to subsequent learning: Early sensory processing is modified to ensure that presentations of similar stimuli are more processed effectively in the future. Thus, this account makes two key experimental predictions, that (*i*) prior knowledge and perceptual learning should affect neural responses in the same brain network and (*ii*) the effect of prior knowledge observed online during perception should predict the magnitude of subsequent perceptual learning. However, because the brain systems supporting the influences of prior knowledge and perceptual learning typically have been observed separately (3, 4, 21–26), neither of these predictions has been tested successfully before.

In this study, we obtained concurrent high-density EEG and magnetoencephalographic (MEG) recordings to compare the

Significance

Experience-dependent changes in sensory processing are critical for successful perception in dynamic and noisy environments. However, the neural and computational mechanisms supporting such changes have remained elusive. Using electrical and magnetic recordings of human brain activity, we demonstrate that two different sources of experience-dependent improvement in perception of degraded speech—the immediate effects of prior knowledge and longer-term changes resulting from perceptual learning—arise from common neural machinery in the sensory cortex although they operate over dramatically different behavioral timescales. Our findings support a specific neuro-computational account of perception in which a single mechanism, minimization of prediction error, drives both the immediate perceptual effects of prior knowledge and longer-term perceptual learning.

Author contributions: E.S. and M.H.D. designed research; E.S. and M.H.D. performed research; E.S. analyzed data; and E.S. and M.H.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] Present address: University College London Ear Institute, London WC1X 8EE, United Kingdom.

[2] To whom correspondence may be addressed. Email: e.sohoglu@gmail.com or matt.davis@mrc-cbu.cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523266113/-/DCSupplemental.

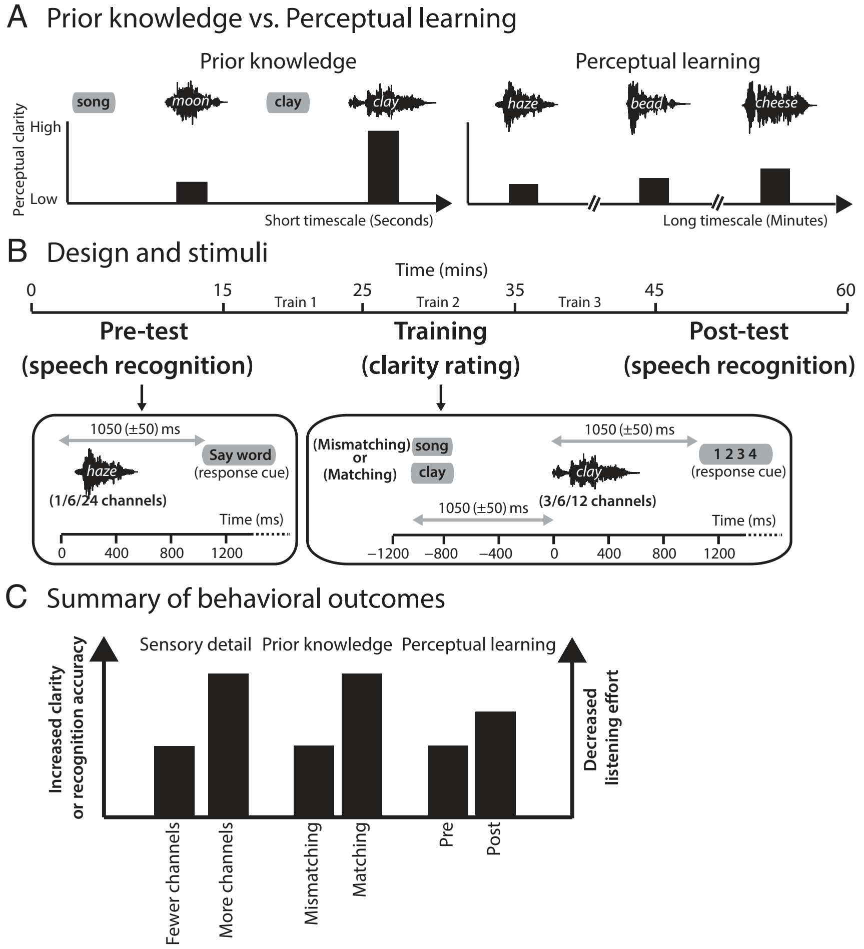

Fig. 1. Overview of study design. (*A*) Illustration of the distinction between the immediate influence of prior knowledge and the more gradual influence of perceptual learning. (*Left*) For mismatching prior knowledge (provided by written presentation of the word "song" before the presentation of the degraded spoken word "moon"), perceptual clarity is low. However, seconds later, perceptual clarity is enhanced dramatically if prior knowledge matches speech content. (*Right*) Perceptual learning results from practice at perceptual tasks (e.g., recognition of degraded speech) leading to gradual improvements in perceptual clarity over a timescale of minutes, hours, or days. (*B*) Timeline of the experiment including a timeline of example trials for the pretest, training, and posttest phases. Training trials were divided into three blocks (Train 1/Train 2/Train 3). (*C*) Summary of changes in behavioral outcomes resulting from experimental manipulations of sensory detail, prior knowledge, and perceptual learning. All three manipulations enhance perceptual clarity or accuracy and decrease listening effort.

impact of prior knowledge and perceptual learning on neural responses to degraded speech (Fig. 1*B*). Speech is an ideal stimulus for exploring this relationship because it is well established that listeners rely heavily on both prior knowledge and perceptual learning for successful perception, especially in noisy conditions or when the speech signal is degraded (6, 27–33). Therefore, using the same stimulus, in the same participants, and in the same experiment, we could test whether these two experience-dependent changes in perception modulate common or dissociable neural mechanisms. In addition, we compared these experience-dependent improvements in speech intelligibility with the improvements resulting from hearing physically clearer speech (Fig. 1*C*). This comparison helps rule out neural changes attributable to changes in listening effort or success that are a downstream consequence of improved perception as opposed to the intrinsic changes in underlying perceptual mechanisms (34).

Results

Behavior. To assess the effect of prior knowledge on the immediate subjective clarity of degraded speech, participants completed a modified version of the clarity-rating task previously used in behavioral and MEG studies (Fig. 1*B*) (23, 35). In this task, listeners are presented with spoken words varying in their amount of sensory detail (and therefore in their intrinsic intelligibility) and are asked to report their subjective experience of speech clarity. Alongside these changes in subjective clarity resulting from physical changes intrinsic to the speech signal, listeners' prior knowledge of the abstract phonological content of speech was manipulated by presenting matching or mismatching text before each spoken word. Consistent with previous findings, and as shown in Fig. 2*A*, speech clarity was enhanced significantly both when sensory detail increased [F (2, 40) = 295, $P < 0.001$] and when listeners had prior knowledge from matching written text [F (1, 20) = 93.2, $P < 0.001$].

This initial period when prior knowledge was used to support perception was designated the "training phase" based on previous work showing that trials in which degraded speech follows matching written text enhance immediate perception (35) and also facilitate longer-term perceptual learning (6, 16). To assess the magnitude of learning in each listener, we measured the accuracy of speech recognition in pretest and posttest evaluations that occurred before and after the training phase (Fig. 1*B*). Participants heard different words across each of the phases of the experiment, enabling us to assess perceptual learning rather than item-specific learning or memory. To ensure further that learning-related changes in perception could be distinguished from other, nonlearning changes (e.g., increased task familiarity, fatigue, and other factors), we included conditions in which we expected speech recognition to be at floor and ceiling levels of accuracy throughout. Specifically, our comparison of post- vs. pretest speech recognition included a more extreme range of sensory detail than obtained during the training phase, i.e., one-channel (unintelligible), six-channel (partially intelligible), and 24-channel (highly intelligible) speech. Perceptual learning would be expected only for the partially intelligible six-channel speech; thus, by assessing the statistical interaction between sensory detail and phase (post- vs. pretest), we could remove any nonlearning influences on behavioral or neural responses. (For a similar approach, see refs. 21 and 22.)

As shown in Fig. 2*B*, we observed a highly significant increase in the accuracy of speech recognition at post- vs. pretest, indicating a robust perceptual learning effect [F (1, 20) = 44.6, $P < 0.001$].

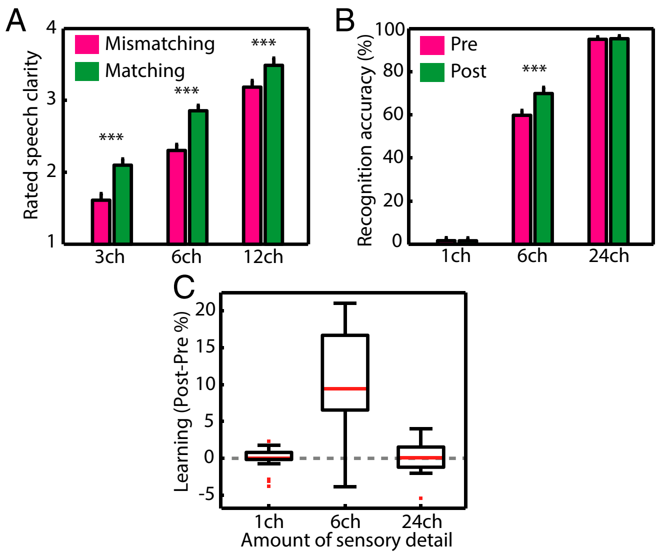

Fig. 2. Group-level behavioral results. (*A*) In the training phase, matching prior knowledge from written text led to an enhancement in immediate speech clarity, as did the provision of increasing speech sensory detail. Error bars represent ± two within-subject SEMs, similar to 95% confidence intervals (85). Asterisks show the significance of planned *t* test comparisons (****P* < 0.001). (*B*) Speech recognition accuracy was enhanced after training (at post-test) compared with before training (at pretest), reflecting long-term perceptual learning. Perceptual learning was significant only for speech with an intermediate amount of sensory detail (six channels). (*C*) Box plots showing variability in the magnitude of improvement in speech recognition (perceptual learning). Red lines indicate the median, black box edges represent 25th and 75th percentiles, black lines cover the range of data points excluding outliers, and red squares mark individual outliers. ch, channels.

Furthermore, there was a significant interaction between phase (pre/posttest) and sensory detail [one, six, or 24 channels; F (2, 40) = 32.8, P < 0.001]. Planned contrasts of posttest > pretest for each level of sensory detail revealed that learning was significant only for six-channel speech [six channels: t (20) = 6.4, P < 0.001; one channel: t (20) = −0.095, P = 0.463; 24 channels: t (20) = 0.543, P = 0.296]. Although the average improvement in the accuracy of speech recognition resulting from learning was ∼10%, individual learning scores varied considerably, with some participants improving by as much as 22% and others showing numerical reductions in recognition accuracy by up to 5% (Fig. 2*C*).

MEG and EEG Training Phase. In the first stage of MEG and EEG analysis, we identified the timing and spatial distribution of neural responses modulated by manipulations of immediate speech clarity during the training phase. We observed effects in a fronto-temporal network for both sensory detail and prior knowledge manipulations that emerged during the late (more than ∼232 ms) portions of the evoked response. All reported effects are family-wise error (FWE)-corrected for multiple comparisons across sensors and time at the P < 0.05 level, unless stated otherwise.

Within the MEG (gradiometer) sensors, the first observable effect of increasing sensory detail involved an increase in neural response at 232–332 ms (Fig. 3*A*). In contrast, providing matching prior knowledge resulted in a decreased MEG response at 232–800 ms (Fig. 3*B*). Distributed source reconstruction using all sensors (EEG and MEG magnetometers and gradiometers) localized both these effects to the left temporal cortex (including the superior temporal gyrus, STG) (Fig. S1*A*). Spatially overlapping but opposite effects of sensory detail and prior knowledge in the STG are consistent with previous observations of neural responses to changes in the subjective clarity of degraded speech (23). As in this previous work, the pattern was reversed in EEG sensors, with decreased and increased neural responses for effects of sensory detail and prior knowledge, respectively (Fig. S2*A*). These latter effects localized to neural sources in the left inferior frontal gyrus and to more posterior frontal sources in pre- and postcentral gyri (Fig. S1*A*). Hence, enhancements in subjective speech clarity modulate late brain responses in a fronto-temporal network; the precise expression of this modulation depends on the source of the clarity enhancement (i.e., increased sensory detail or matching prior knowledge) and the neural locus of underlying generators (i.e., temporal or frontal cortex). Note that if these effects were caused simply by reduced listening effort rather than by more specific changes in intrinsic perceptual processing, one would not expect to observe any of the above dissociations between the manipulations of sensory detail and prior knowledge. We will return to this point in *Discussion*.

For our next stage of analysis, we tested whether neural processes that are modulated by prior knowledge also contribute to longer-term learning (as assessed by improved recognition accuracy at post- vs. pretest). We used the sensor-space clusters showing a significant effect of prior knowledge to define sensor × time volumes of interest (VOIs) from which we extracted the MEG/EEG signal (averaged across sensors and time for each participant). We then tested whether the difference between matching and mismatching conditions in each of these VOIs was significantly correlated across subjects with the magnitude of improvement in

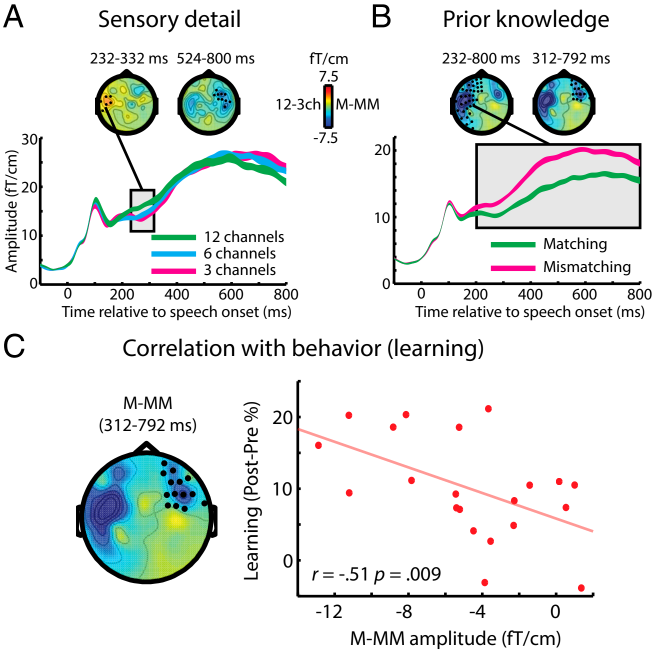

Fig. 3. Group-level effects in MEG (gradiometer) sensors during the training phase. (*A*) Main effect of sensory detail on sensor-space data. Topographic plots show statistical clusters showing a main effect of sensory detail (one plot for each cluster). Black circles within and the text directly above each scalp topography indicate the spatial and temporal extent of each cluster, respectively. The topographies themselves represent the difference in MEG response between 12-channel and three-channel speech, averaged over the temporal extent of the clusters. Waveforms represent the MEG response averaged over the spatial extent of a single selected cluster (indicated by the connecting black line). The width of waveforms represents ± two within-subject SEMs, similar to 95% confidence intervals. (*B*) The main effect of prior knowledge on sensor-space data. Topographies represent the difference in MEG response to speech presented after matching or mismatching written text, averaged over the temporal extent of significant clusters. (*C*) Right-hemisphere MEG cluster (at 312–792 ms) within which the neural effect of prior knowledge (matching/mismatching) correlated across subjects with the behavioral magnitude of perceptual learning of six-channel speech (speech recognition improvement post- vs. pretest). ch, channels; M, matching; MM, mismatching.

speech recognition post- vs. pretest (measured in the six-channel condition for which performance was not at floor or ceiling). Note that these two dependent measures are statistically independent (36); therefore we could test for correlations between them without double-dipping (37, 38). As shown in Fig. 3*C*, only one MEG VOI (in the right hemisphere, at 312–792 ms) that was modulated by prior knowledge significantly predicted individual differences in the magnitude of learning [one-tailed Pearson's $r = -0.51$, $n = 21$, P (uncorrected) = 0.009; $P < 0.025$ Bonferroni-corrected across two MEG VOIs].

MEG and EEG Pre- and Posttest Phases. Having characterized the manner in which neural responses were modulated by prior knowledge during training, we next assessed how neural responses changed from the pre- to the posttest phase; these neural changes accompany enhanced speech perception resulting from perceptual learning. As in the training phase, speech was presented with different amounts of sensory detail. Hence, we first tested whether the same effect of sensory detail was present as during the training phase, despite listeners performing a different (speech recognition) task. As shown in Fig. 4*A*, increasing sensory detail again resulted in an increased MEG response at 200–800 ms, similar to our previous MEG observations for this manipulation during clarity-rating tasks (23) and for hemodynamic studies that used either clarity-rating (39, 40) or speech-recognition tasks (22, 24, 41).

We next assessed the effects of our critical learning manipulation (i.e., the changes in neural responses post- vs. pretest). As shown in Fig. 4*B*, the first observable effects of learning occurred at 40 ms after speech onset and involved reductions in MEG responses following training (one cluster over the left-hemisphere sensors at 40–112 ms and another over the right hemisphere at 68–108 ms). A later effect also was present at 448–760 ms but with an opposite polarity (i.e., increased MEG response posttraining).

To test whether neural responses post- vs. pretest reflect perceptual learning, as opposed to possible confounding factors such as task familiarity, we assessed the statistical interaction between experiment phase (pre/post test) and sensory detail (one, six, or 24 channels). We tested for this interaction in VOIs defined from the clusters showing a significant effect of post- vs. pretest (similar to our previous VOI analysis for the training phase). As shown in Fig. 4*C*, the early post- vs. pretest VOI in the right hemisphere (at 68–108 ms) showed a significant interaction with sensory detail [F (2, 40) = 6.63, P (uncorrected) = 0.004; $P < 0.025$ Bonferroni-corrected across three MEG VOIs]. The same interaction effect was apparent in the left hemisphere (at 40–112 ms), at an uncorrected threshold [F (2, 40) = 3.33, P (uncorrected) = 0.046; $P = 0.138$ Bonferroni-corrected across three MEG VOIs]. In the right-hemisphere VOI (68–108 ms), where the interaction was more reliable, pairwise post hoc comparisons for each sensory detail condition revealed a significant decrease in MEG signal at post- vs. pretest for six-channel speech, as observed in the behavioral data [one-tailed t (20) = −6.99, $P < 10^{-6}$; $P < 0.001$ Bonferroni-corrected across three comparisons] and also for 24-channel speech [one-tailed t (20) = −5.26, $P < 10^{-4}$; $P < 0.001$ Bonferroni-corrected across three comparisons]. In contrast, there was no significant decrease for one-channel speech [one-tailed t (20) = −1.26, $P = 0.112$; $P = 0.335$ Bonferroni-corrected across three comparisons]. Distributed source reconstruction of the interaction effect {[posttest – pretest (six channels + 24 channels)] < [posttest – pretest (one channel)]} revealed a bilateral neural source in the STG (Fig. S1*B*).

Thus, like the effect of prior knowledge, learning modulated neural responses in the STG. Intriguingly, these two effects occurred at markedly different latencies (>232 ms for prior knowledge during training vs. 68–108 ms for perceptual learning shown in the comparison of post- vs. pretest), a point to which we will return in *Discussion*. Additionally, the observation of post- vs. pretest changes in early MEG responses for 24-channel (as well as for six-channel) speech suggests that the neural effects of perceptual learning occur whenever speech is degraded but intelligible, whether or not behavioral changes in recognition accuracy are observed (accuracy scores were at ceiling for 24-channel speech

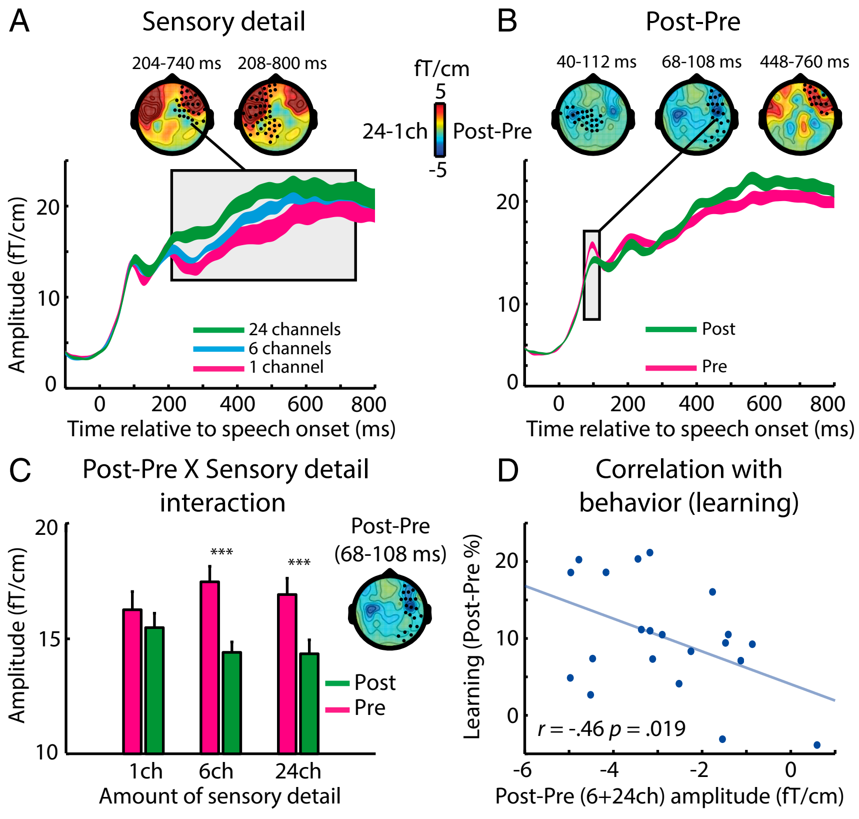

Fig. 4. Group-level effects in MEG (gradiometer) sensors during pre- and posttest phases. Data are plotted as in Fig. 3. (*A*) Main effect of sensory detail on sensor-space data. The topographies represent the difference in MEG response between 24-channel and one-channel speech. (*B*) Main effect of perceptual learning (post- vs. pretest) on sensor-space data. The topographies represent the difference in MEG/EEG response between post- and pretest phases. (*C*) In the right-hemisphere MEG cluster (68–108 ms) the neural reduction resulting from perceptual learning (post- vs. pretest) depended on the amount of speech sensory detail. Error bars indicate ± two within-subject SEMs. Asterisks show the significance of post hoc t tests, Bonferroni-corrected for multiple comparisons (***$P < 0.001$). (*D*) The neural effect of perceptual learning (post- vs. pretest averaged over six- and 24-channel speech) correlated across subjects with the magnitude of learning-related change in behavior (improvement in speech recognition post- vs. pretest for six-channel speech). ch, channels.

throughout). Behavioral measures that are more finely grained than recognition accuracy, such as response times or confidence, might reveal a behavioral effect for the 24-channel condition also. Note that our design is not optimal for observing the effects of response time, because listeners were asked to wait for a cue before giving their response to avoid contaminating the MEG data with motor responses. Nonetheless, the presence of an interaction between sensory detail and post- vs. pretest allows us to rule out explanations of this effect in terms of nonlearning influences (e.g., increased task familiarity, fatigue, and other factors) on neural responses.

As an additional test of whether early MEG differences between responses post- vs. pretest were caused by neural effects of perceptual learning, we correlated the MEG signal in each of the two VOIs described above (averaged over the six- and 24-channel conditions that showed the strongest neural reductions) with the magnitude of improvement in recognition accuracy post- vs. pretest (for six-channel speech only, because performance was at ceiling for 24-channel speech). As shown in Fig. 4*D*, we found a significant correlation across subjects in the right hemisphere VOI [one-tailed Pearson's $r = -0.46$, $n = 21$, P (uncorrected) $= 0.019$; $P < 0.05$ Bonferroni-corrected across two MEG VOIs]. This test also was significant when correlating the MEG signal with the improvement in recognition accuracy averaged over the six- and 24-channel conditions [one-tailed Pearson's $r = -0.51$, $n = 21$, P (uncorrected) $= 0.0087$; $P < 0.025$ Bonferroni corrected across two MEG VOIs]. In the left hemisphere, where the interaction between phase and sensory detail was weaker, there was no corresponding correlation (one-tailed Pearson's $r = -0.042$).

EEG sensor-space responses for this phase of the experiment were analyzed also and showed a similar reduction post- vs. pretest, although at an intermediate latency of 124–220 ms (Fig. S2*B*). However, this effect did not interact with sensory detail, nor did it correlate with the magnitude of improvement in speech recognition over the test phases.

A final test for this stage of our analysis was designed to determine whether the early effect of perceptual learning on neural responses post- vs. pretest also was present during the training phase. We used the two early clusters showing an effect of post- vs. pretest to define VOIs within which we assessed how MEG responses evolved over the three blocks of the training phase. As with the early effect of post- vs. pretest, the right hemisphere VOI (at 68–108 ms) showed a significantly reduced MEG response over the three training blocks [F (2, 40) $= 7.42$, P (uncorrected) $= 0.0018$; $P < 0.05$ Bonferroni-corrected across two MEG VOIs] (Fig. S3*A*). Furthermore, planned comparisons revealed a significant correlation across subjects between the train 3 < train 1 effect and the magnitude of improvement in the recognition of six-channel speech post- vs. pretest (one-tailed Pearson's $r = -0.54$, $n = 21$, $P < 0.01$) (Fig. S3*B*). In the left-hemisphere VOI (at 40–112 ms), the MEG response also was reduced significantly over the three training blocks [F (2, 40) $= 11.80$, $P < 0.001$], but the contrast of train 3 < train 1 did not correlate significantly across subjects with the magnitude of improvement in speech recognition post- vs. pretest (one-tailed Pearson's $r = -0.24$).

Source Dipole Analysis. In our final analysis, we further tested whether the same neural source in the temporal lobe was modulated by prior knowledge and perceptual learning or whether these two effects originated from spatially distinct neural sources (e.g., in the STG versus the middle temporal gyrus) (25, 26). We did so by using a more constrained method of source reconstruction (42) in which the center of neural activity in a local cortical patch was modeled as a single focal source (an equivalent current dipole, ECD) with a "soft" Bayesian prior for locations in bilateral STG.

We used two ECD sources (one in each hemisphere) to model single-subject MEG activity modulated by prior knowledge at 312–792 ms (matching–mismatching) and by perceptual learning at 68–108 ms {[post-pre(6+24 channels)] − [post-pre(one-channel)]}. As shown in Fig. 5*A*, the mean locations for both prior knowledge (left hemisphere: $x = -46$, $y = -29$, $z = +5$; right hemisphere: $x = +44$, $y = -27$, $z = +6$) and perceptual learning (left hemisphere: $x = -47$, $y = -26$, $z = +12$; right hemisphere: $x = +53$, $y = -26$, $z = +3$) manipulations were estimated to lie within the STG [assigned objectively by the Statistical Parametric Mapping (SPM) anatomy toolbox; see ref. 43] and in close proximity to each other (mean locations were 7.68 mm apart in the left hemisphere, and 9.54 mm apart in the right hemisphere). Repeated-measures ANOVA of MNI coordinates with manipulation (prior knowledge/perceptual learning) and hemisphere (left/right) as factors revealed a significant interaction between manipulation and hemisphere along the superior/inferior axis [F (1, 20) $= 5.49$, $P < 0.05$]. Planned contrasts revealed a more superior STG source modulated by perceptual learning vs. prior knowledge only in the left hemisphere [left hemisphere: one-tailed t (20) $= 2.06$, $P < 0.05$; right hemisphere: one-tailed t (20) $= 0.538$, $P < 0.3$]. No other differences in location were significant, although there was a marginal main effect of manipulation on distance from the midline [F (1, 20) $= 4.25$, $P = 0.053$] together with a marginal interaction of manipulation × hemisphere [F (1, 20) $= 3.07$, $P = 0.095$] reflecting a tendency for a more lateral source modulated by prior knowledge in the right hemisphere [right hemisphere: one-tailed t (20) $= 2.76$, $P = 0.006$; left hemisphere: one-tailed t (20) $= 0.212$, $P = 0.418$]. The same pattern of results was obtained when looser constraints on the prior locations were used (*Methods*).

This ECD analysis therefore suggests that the prior knowledge and perceptual learning effects have nearly identical spatial origins. Although the precise locations are statistically different in the left hemisphere, both locations were confirmed to reside in the same anatomical structure (left STG). The corresponding locations in the right hemisphere are numerically indistinguishable. The difference in ECD location observed in the left hemisphere could reflect a subtle difference in the distributions of neural activity occurring within overlapping regions of the STG, leading to differently estimated locations for the modeled sources (44).

Planned contrasts of source strength support earlier sensor-space and distributed-source reconstruction analyses (Fig. 5*B*). Bilateral STG showed significant reductions in neural activity resulting from prior knowledge [at 312 to 792 ms; left hemisphere: t (20) $= -8.58$, $P < 0.001$; right hemisphere: t (20) $= -2.68$, $P < 0.01$] and from perceptual learning [at 68 to 108 ms; left hemisphere: t (20) $= -2.87$, $P < 0.01$; right hemisphere: t (20) $= -3.28$, $P < 0.01$]. We also confirmed that increasing sensory detail (24 channels vs. one channel) led to the opposite effect, i.e., increased activity, in this region [from 204 to 740 ms; left hemisphere: t (20) $=$ 5.58, $P < 0.001$; right hemisphere: t (20) $= 3.43$, $P < 0.01$]. Furthermore, as shown in Fig. 5 *C* and *D*, the reduction in the magnitude of source strength in the right STG resulting from prior knowledge (matching vs. mismatching) correlated significantly across subjects with the behavioral magnitude of six-channel speech learning (one-tailed Pearson's $r = -0.40$, $n = 21$, $P < 0.05$), as did the reduction in the magnitude of activity in right STG resulting from perceptual learning (post- vs. pretest averaged over six and 24 channels; one-tailed Pearson's $r = -0.55$, $n = 21$, $P < 0.01$). As with earlier sensor-space analyses, these correlations were not present in the left hemisphere (prior knowledge: one-tailed Pearson's $r = -0.09$; perceptual learning: one-tailed Pearson's $r = -0.09$).

Discussion

In the current study, we tested whether the influences of prior knowledge and perceptual learning on speech perception arise through common or dissociable neural mechanisms. When the perception of degraded speech was supported by prior knowledge of abstract phonological content (provided by the prior presentation of matching written text), we replicated previous findings of reduced speech-evoked responses in a peri-auditory region of the STG (23). Critically, perceptual learning reduced activity in a nearly identical region of the STG, and this reduction correlated across subjects not only with the magnitude of the learning-related change in behavior but also with the magnitude

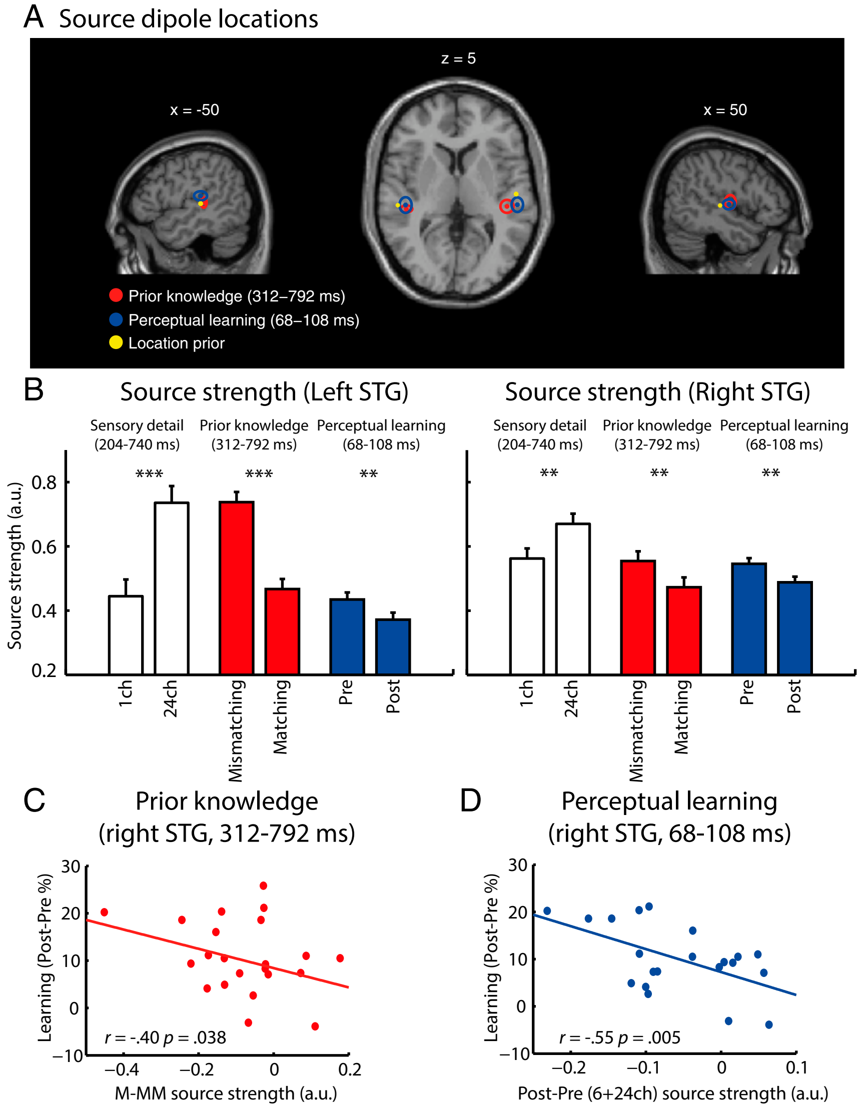

Fig. 5. Group-level source dipole analysis of MEG gradiometer data. (*A*) Group means of source dipole locations (shown as circles) overlaid onto an MNI-space template brain for the neural effect of prior knowledge (matching–mismatching) and perceptual learning {[post–pretest (6+24 channels)] < [post–pretest (one channel)]}. The spatial extent of the ellipses represents ± two SEMs. Also shown (yellow circle) is the location prior mean used for the dipole estimation procedure. (*B*) Dipole source strength in the left and right STG as a function of increasing sensory detail (24 channels vs. one channel), prior knowledge (matching–mismatching), and perceptual learning (post- vs. pretests). Error bars indicate ± two within-subject SEMs. Asterisks show the significance of planned comparisons: **$P < 0.01$; ***$P < 0.001$. (*C*) Right STG dipole within which the effect of prior knowledge (matching–mismatching conditions) on source strength correlated across subjects with the behavioral magnitude of perceptual learning (improvement in speech recognition post- vs. pretest for six-channel speech). (*D*) Right STG dipole within which the effect of perceptual learning (post- vs. pretest averaged over six- and 24-channel speech) on source strength correlated across subjects with the magnitude of learning-related change in behavior (improvement in speech recognition post- vs. pretest for six-channel speech). ch, channels; M, matching; MM, mismatching.

of the reduction in STG activity caused by prior knowledge. Thus, these findings are consistent with a single-mechanism account: More accurate prior expectations for degraded speech drive both immediate and longer-term changes in sensory processing.

Common Mechanisms for Prior Knowledge and Perceptual Learning. Previous behavioral studies have shown that the provision of relevant prior knowledge enhances both immediate perceptual clarity (1, 29, 35, 45, 46) and perceptual learning of degraded speech (6, 14, 16–18). However, our work goes beyond these behavioral studies by identifying a common neural signal (reductions in the STG response) that is associated with both these effects. Thus, we present the strongest evidence to date for a single underlying mechanism.

One key issue in proposing a single-mechanism account is ruling out the possibility that reductions in neural activity caused by prior expectations and perceptual learning reflect changes in listening effort that are a downstream consequence of enhanced sensory processing. The need to distinguish between changes that reflect the outcome of learning and the mechanisms that support learning itself is an acknowledged issue for studies of perceptual learning in a range of domains (34). The present study dissociated the neural effects of prior knowledge and perceptual learning from effects caused by changes in sensory detail. Rather than reducing neural responses, improved speech intelligibility resulting from physically clearer speech produced an increased STG response (compare Figs. 1*C* and 5*B*). This increased response rules out the possibility that reductions in response reductions are caused by epiphenomenal changes in listening effort. Instead, our results support more specific conclusions concerning predictive neural mechanisms that we argue are responsible for changes in perception; we expand on these conclusions later.

Despite the cross-subject correlations between the influences of prior knowledge and perceptual learning, with nearly identical spatial origins, we observed differences in the timing of immediate and longer-term neural changes. The perceptual learning effect in the STG was revealed as an early (~100 ms) reduction in the magnitude of the speech-evoked response, whereas the response reduction caused by matching prior knowledge was observed at a later latency (from 232 ms onward) as a sustained modulation of the speech-evoked response. Despite this difference in timing, we maintain that these findings are consistent with single-mechanism accounts. In these single-mechanism accounts, the influence of prior knowledge on sensory processing is contingent on information returning from higher-order levels of representation (19, 20). In our study, these higher-level representations were derived from prior written text, which can provide only abstract (i.e., nonacoustic) information about the phonological content in speech (23, 47). The processing of this higher-level information may invoke a delay in the top-down effect of prior knowledge on sensory processing in the STG (e.g., if the higher-level phonological correspondence between speech and prior written text is first assessed by frontal or somatomotor regions). However, once top-down modulation of STG activity has occurred, any effect will be preserved in the long-term changes in synaptic connectivity intrinsic to the STG or in the connectivity between this region and higher-level cortex. In subsequent trials, these long-term changes will shape early, bottom-up processing of future speech signals, even in the absence of top-down feedback (20).

The timing of the response reduction caused by perceptual learning in the STG is remarkably early. This finding is without precedent in previous functional MRI studies investigating perceptual learning of degraded (vocoded) speech, not only because of the lack of temporal resolution in previous studies but also because these studies primarily have observed the effects of learning in the thalamic or parietal and prefrontal regions (21, 22) but not in the STG regions that are central to functional neuroanatomical accounts of speech perception (48, 49). However, the finding is consistent with other neurophysiological and eye-tracking studies showing early changes in speech processing after perceptual learning of synthetic (50–52) and ambiguous (53) speech sounds and suggests that the long-term changes in sensory processing observed here might have a more general role in supporting the perception of degraded and accented speech. In the next paragraphs we discuss some of the implications of these findings for a single-mechanism account of perception and perceptual learning of speech based on predictive coding principles.

A Predictive Coding Account of Speech Perception and Perceptual Learning. Several computational accounts have proposed that a single mechanism underpins the influences of both prior knowledge and perceptual learning (19, 20, 54). We argue that one account in particular, predictive coding, explains our findings most completely (19, 55–61). This account, depicted in Fig. 6*A*, proposes that perception arises from a hierarchical Bayesian inference process in which prior knowledge is used top-down to predict sensory representations at lower levels of the cortical hierarchy. The difference between the observed and predicted sensory input is computed at lower levels, and only discrepant sensory input (prediction error) is propagated forward to update higher-level perceptual interpretations. This predictive coding theory makes a specific neural proposal concerning the mechanisms by which the immediate influences of prior knowledge and longer-term perceptual learning operate. In this view, each time neural processing changes online to reduce prediction error (e.g., during exposure to spoken words preceded by matching written text), synaptic connectivity between higher-level and sensory representations (and vice-versa) is adjusted incrementally to match the long-term statistics of sensory input, thereby also reducing prediction error for future presentations of similar sounds (e.g., speech produced with similar vocoding parameters or by talkers with the same accent).

Computational simulations illustrate how this predictive coding architecture provides a unifying explanation for the observed dissociation between the neural effect of increased sensory detail and the effects of prior knowledge and perceptual learning; a summary of the simulation method is shown in Fig. 6*A* and simulation outcomes are shown in Fig. 6 *B* and *C*; full details are presented in Fig. S4 and in *SI Methods* and simulation parameters are listed in Table S1. These simulations show how listening conditions in which top-down predictions can more accurately explain sensory input (e.g., from prior knowledge or perceptual learning) result in reduced prediction error and therefore in reduced neural activity while improving perceptual outcomes. Critically, this predictive coding mechanism explains how increases in sensory detail lead to the opposite effect on neural responses, despite producing the same behavioral outcome: An increase in sensory information necessarily results in a larger prediction error unless there is an accompanying prediction for that sensory information. These opposing effects are consistent with our observations in the STG and are difficult to explain with accounts in which STG activity is a simple function of perceptual clarity or listening effort (23). Thus, our results are better explained by predictive coding theory in which a single mechanism, minimization of prediction error, drives the immediate perceptual effects of prior knowledge and sensory detail as well as the longer-term perceptual learning of degraded speech.

To simulate longer-term perceptual learning of degraded speech, we use changes in the precision or variance of sensory predictions. Optimal perceptual outcomes and minimal prediction errors occur when the precision of sensory predictions match the precision of the sensory input (62). Such changes increase the amount of information in updated perceptual hypotheses and hence the accuracy of perceptual outcomes while decreasing the magnitude of prediction error. These changes therefore are in line with the behavioral and neural effects of perceptual learning observed in our experiment. As described in *SI Discussion*, this account provides a neural implementation of "attentional weighting" theories of perceptual learning (5); with learning arising from Hebbian weight updates that minimize prediction errors when degraded speech matches prior predictions. This account therefore explains why perceptual learning of vocoded speech is enhanced when the content of degraded speech is predicted accurately (6, 18); these trials lead to learning by allowing listeners to attend more appropriately to informative sensory features in degraded speech (5, 7, 62, 63).

Methods

Participants. Twenty-one (12 female, 9 male) right-handed participants were tested after giving informed consent under a process approved by the Cambridge Psychology Research Ethics Committee. All were native English speakers, aged 18–40 y (mean ± SD, 22 ± 2 y) and had no history of hearing impairment or neurological disease based on self-report.

Spoken Stimuli. A total of 936 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 372–903 ms (mean ± SD = 591 ± 78 ms).

The amount of sensory detail in speech was varied using a noise-vocoding procedure (64), which superimposes the temporal envelope from separate frequency regions in the speech signal onto white noise filtered into corresponding frequency regions. This procedure allows parametric variation of spectral detail, with increasing numbers of channels associated with increasing intelligibility. Vocoding was performed using a custom Matlab script (The MathWorks, Inc.), using 1, 3, 6, 12, or 24 spectral channels logarithmically spaced between 70 and 5,000 Hz. 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.

Each spoken word was presented only once in the experiment so that unique words were heard in all trials. The particular words assigned to each condition were randomized across participants. Before starting the speech-recognition and clarity-rating tasks, participants completed brief practice sessions, each lasting approximately 5 min, that contained all the conditions of the subsequent experimental phase but used a corpus of words different from those used in the main experiment. Stimulus delivery was controlled with E-Prime 2.0 software (Psychology Software Tools, Inc.).

Training Phase. During the training phase, participants completed a modified version of the clarity-rating task previously used in behavioral and MEG studies combined with a manipulation of prior knowledge previously shown to enhance perceptual learning of degraded (vocoded) speech (Fig. 1*B*) (6, 18, 23). Speech was presented with three, six, or 12 channels of sensory detail. Prior knowledge of speech content was manipulated by presenting mismatching or matching text before speech onset (Fig. 1*B*). Written text was composed of black lowercase characters presented for 200 ms on a gray background. Mismatching text was obtained by permuting the word list for the spoken words. As a result, each written word in the mismatching condition also was presented as a spoken word in a previous or subsequent trial, and vice versa.

Trials commenced with the presentation of a written word, followed 1,050 (±0–50) ms later by the presentation of a spoken word (Fig. 1*B*). Participants were cued to respond by rating the clarity of each spoken word on a scale from 1 (not clear) to 4 (very clear) at 1,050 (±0–50) ms after speech onset. The response cue consisted of a visual display of the rating scale, and responses

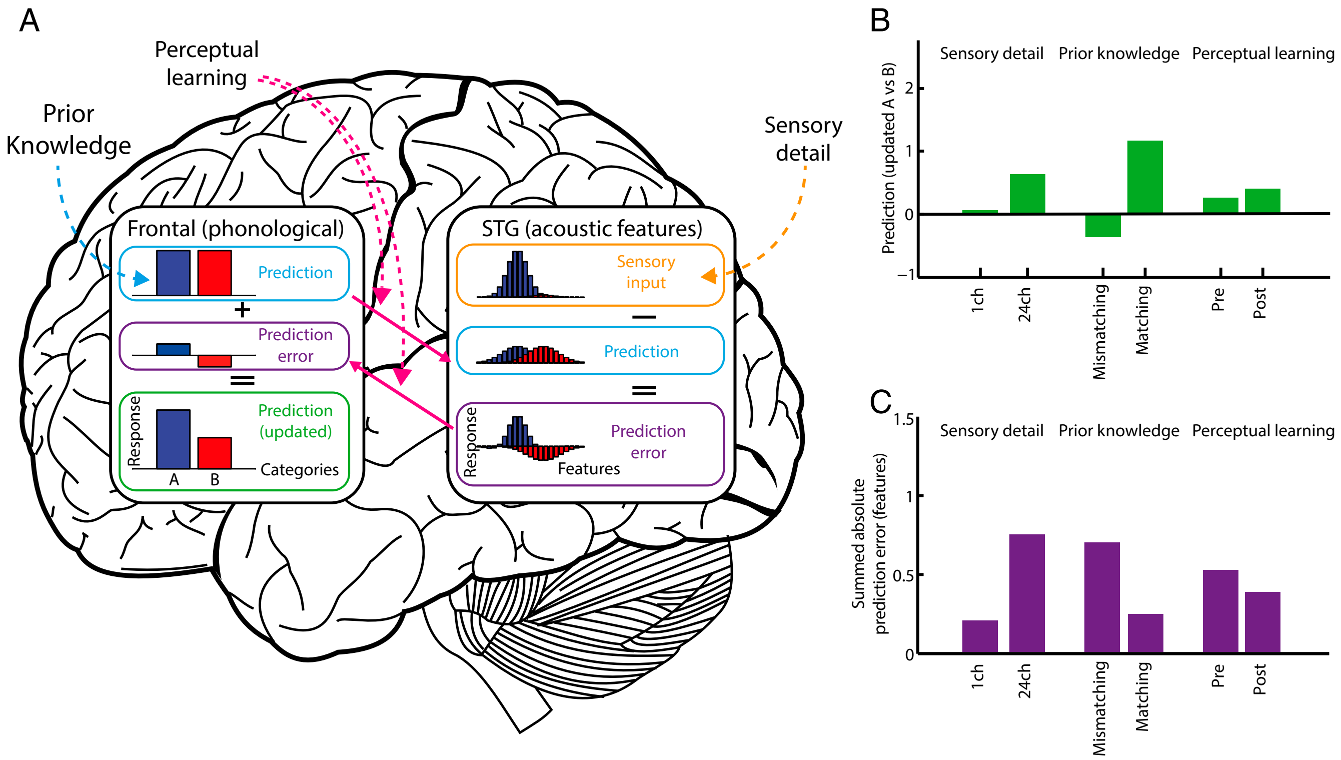

Fig. 6. Depiction of a predictive coding theory of the perception of degraded speech based on the current data, following standard assumptions (19, 55–61). (*A*) The graphs depict simulated responses when hearing 24-channel vocoded speech during pre- and posttest phases (i.e., without a matching/mismatching prior). For responses in other conditions see Fig. S4; for simulation methods see *SI Methods* and Table S1. Behavioral and neural outcomes are determined by the interactions between two hierarchically organized levels of representation: sensory (acoustic–phonetic) features (STG) and phonological categories (frontal and somatomotor regions). These representations are depicted in the bar graphs, intended to reflect neuronal activation over different regions of the cortex and organized by perceptual similarity. Perceptual hypotheses about the phonological content of speech (e.g., two phonological categories, A and B, color-coded blue and red, respectively) are formed in frontal regions and conveyed top-down to the STG as predictions for upcoming sensory features via weights (solid pink arrow) that encode expected feature values (e.g., voice onset time, VOT). We have color coded these phonological predictions (and the corresponding sensory feature predictions and prediction errors) to indicate their associated category, although activation values at the sensory level are always summed over the two categories. These feature predictions are subtracted from the pattern of sensory input received by the STG, and the resulting feature prediction errors then are returned via weighted connections (solid pink arrow) to the frontal regions and are used to update predictions for categories A and B. In this scheme, increases in speech sensory detail (broken orange arrow) produce more informative feature-activation patterns in sensory input units that favor certain phonological categories (in the depicted example, a sensory signal for category A). In contrast, both prior knowledge and perceptual learning manipulations modulate top-down predictions for future sensory activation patterns, although in different ways. Changes in prior knowledge (broken blue arrow) modulate the relative likelihood of the two categories encoded by activity in phonological prediction units. Perceptual learning (broken pink arrows) modulates the connection weights that map between categories and features. (*B*) Simulated perceptual outcomes based on updated predictions in frontal phonological units for key experimental conditions. The relative magnitude of the updated phonological predictions for categories A versus B [log(A)-log(B)] accurately simulates the qualitative pattern observed in behavioral data: Clarity ratings and recognition accuracy for category A are increased by all three manipulations (comparable to the changes shown in Fig. 1*C*). (*C*) Simulated feature prediction error values in different perceptual conditions. The summed (absolute) prediction error simulates the qualitative pattern of neural responses in the STG (Fig. 5*B*). ch, channels.

were recorded by a four-button box manipulated by the participant's right hand. Subsequent trials began 850 (±0–50) ms after the participant responded.

Manipulations of sensory detail (three-, six-, or 12-channel speech) and prior knowledge of speech content (mismatching/matching) were fully crossed, resulting in a 3 × 2 factorial design for this part of the experiment with 78 trials in each condition. Trials were randomly ordered during each of three presentation blocks of 156 trials.

Pre- and Posttest Phases. Before and after the training phase, participants completed a speech-recognition task using unintelligible, partially intelligible, or highly intelligible vocoded words. Speech was presented with one, six, or 24 channels of sensory detail, and participants were cued to respond by reporting each spoken word 1,050 (±0–50) ms after speech onset (Fig. 1*B*). The response cue consisted of a visual display of the words "Say word," and vocal responses were recorded with a microphone. Subsequent trials began 850 (±0–50) ms after the participant pressed a button on a response box using the right hand.

To calculate the accuracy of speech recognition, vocal responses were first transcribed using the DISC phonemic transcription in the CELEX database (65). These transcriptions subsequently were compared with the phonemic transcriptions of the word stimuli using a Levenshtein distance metric that measures the dissimilarity between two strings (66). To convert this metric into a measure of speech-recognition accuracy, the Levenshtein distance for each stimulus–response pair was expressed as a percentage of the length of the longest string in the pair and was subtracted from 100%. The result is a highly sensitive measure of speech-recognition accuracy, similar to the proportion of segments correctly recognized but with partial credit given for segments recognized correctly but in incorrect positions. Even words that are incorrectly recognized can produce highly accurate scores (e.g., reporting "haze" as "daze" would result in a score of 67%).

For this part of the experiment sensory-detail conditions (one-, six-, or 24-channel speech) and phase (pre/post test) were combined to produce a 3 × 2 factorial design with 78 trials in each condition. The pre- and posttest phases each consisted of 234 randomly ordered trials sampled equally from the three sensory-detail conditions.

Data Acquisition and Preprocessing. Magnetic fields were recorded with a VectorView system (Elekta Neuromag) containing a magnetometer and two orthogonal planar gradiometers at each of 102 positions within a hemispheric array. Electric potentials were recorded simultaneously using 68 Ag-AgCl sensors according to the extended 10–10% system and referenced to a sensor placed on the participant's nose. All data were digitally sampled at 1 kHz and high-pass filtered above 0.01 Hz. Head position and electro-oculography activity were monitored continuously using four head-position indicator (HPI) coils and two bipolar electrodes, respectively. A 3D digitizer (FASTRAK; Polhemus, Inc.) 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 preauricular points).

Data from the MEG sensors (magnetometers and gradiometers) were processed using the temporal extension of Signal Source Separation (67) in MaxFilter software (Elekta Neuromag) to suppress noise sources, compensate for motion, and reconstruct any bad sensors. Noisy EEG sensors were identified by visual inspection and were excluded from further analysis. Subsequent processing was done in SPM8 (Wellcome Trust Centre for Neuroimaging) and FieldTrip (Donders Institute for Brain, Cognition and Behavior) software implemented in Matlab. The data were down-sampled to 250 Hz and epoched from −100 to 800 ms relative to speech onset. After epoching, the data were baseline-corrected relative to the 100-ms prespeech period and low-pass filtered below 40 Hz, and the EEG data were referenced to the average over all EEG sensors. Finally, the data were robust averaged across trials (68, 69) to down-weight outlying samples, minimize non–phase-locked activity, and derive the evoked response. To remove any high-frequency components that were introduced to the data by the robust averaging procedure, low-pass filtering was repeated after averaging.

Sensor-Space Statistical Analysis. Before statistical analysis, the data were converted into 3D (2D sensor × time) images by spherically projecting onto a 32 × 32 pixel plane for each epoch time-sample (between 0 and 800 ms) and were smoothed using a 10 mm × 10 mm × 25 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. Following conversion into images, *F* tests for main effects were performed across sensors and time while controlling the FWE rate across all three data dimensions using random field theory (70). Reported effects were obtained by using a cluster defining a height threshold of $P < 0.001$ with a cluster extent threshold of $P < 0.05$ (FWE-corrected) under nonstationary assumptions (71).

Follow-up interactions and correlations with behavior were conducted on MEG/EEG signals averaged across sensors and time from clusters showing significant main effects (see *Results* for more details). Importantly, these follow-up tests are statistically independent from the main effects and hence can be conducted without double-dipping (36–38).

Source Reconstruction. To determine the underlying brain sources of the sensor-space effects, distributed models were first used to reconstruct activity across the whole cortex. The results of this first-source reconstruction subsequently informed a more constrained ECD method of reconstructing focal sources. Both these analyses depended on participant-specific forward models, using single shell and boundary element models for the MEG and EEG sensors, respectively. Computation of these forward models involved the spatial normalization of a T1-weighted structural MRI scan obtained from each participant to the MNI template brain in SPM8. The inverse transform of this spatial normalization was used to warp the cortical, inner skull, outer skull, and scalp meshes of the template brain to the participant's MRI space. Sensor positions were projected onto each subject's MRI space by minimizing the sum of squared differences between the digitized fiducials and the MRI scan fiducials and between the digitized head shape and the template scalp mesh. For five participants, a structural MRI scan was not available; in these cases the spatial normalization of the MNI template brain was based on the digitized fiducials and head shape.

For the distributed models, a multimodal source inversion scheme was used to integrate data from all three neurophysiological measurement modalities (EEG and MEG magnetometers and gradiometers); this scheme has been shown to give more precise localization than obtained by considering each modality in isolation (72). This scheme is implemented within the parametric empirical Bayes framework of SPM8 (73–75). This approach allows the use of multiple priors, in the form of source covariance matrices, which constrain the resulting source solutions and are optimized by maximizing the negative free-energy approximation to the model evidence (76). In the current study, we used the LOR set of priors in SPM8, so that all sources were assumed a priori to be equally activated (by specifying the source covariance matrix as an identity matrix) and spatially correlated (by introducing a spatial dependency between mesh vertices that were, on average, 6 mm apart). This method produces smooth solutions similar to those obtained by the LORETA method (77), which are appropriate for assessing activation overlap across participants. Multimodal fusions of the data were achieved by using a heuristic to convert all sensor data to a common scale and by weighting additional error covariance matrices for each sensor type to maximize the model evidence (72, 75). Such an approach allows the noise levels associated with each sensor type to be estimated directly from the data; by maximizing the negative free energy, sensor types with high estimated levels of noise contribute less to the resulting source solutions. A final constraint was imposed so that source solutions were consistent across participants; this constraint has been shown to improve group-level statistical power (74). Significant effects from sensor space were localized within the brain by summarizing source power in the 1–40 Hz range for each participant and the time window of interest using a Morlet wavelet projector (78). Time windows were selected based on the temporal extent of significant statistical clusters observed in sensor space. One exception to this method is the temporally extended increase in EEG response resulting from matching prior knowledge that occurs 148–800 ms after speech onset (Figs. S1*A* and S2*A*). Because source reconstruction during this extended time window did not reveal reliable activations, we shortened the time window to 148–232 ms to localize only the onset of this prior-knowledge effect. Source power estimates subsequently were converted into one-tailed pairwise *t* statistics for each effect of interest. Given that the goal of source reconstruction was to localize the neural generators of sensor-space effects previously identified as significant, statistical maps of source activity are displayed with an uncorrected voxelwise threshold ($P < 0.05$).

Two ECD models were computed for each participant's MEG planar gradiometer data using the variational Bayes scheme implemented within SPM8 (42): one for the effect of prior knowledge time-averaged at 312–792 ms and another for the effect of perceptual learning at 68–108 ms. To avoid local maxima, the ECD procedure was run 100 times for each model using different initial location and moment parameters and selecting the solution with the highest model evidence. The mean prior locations for these models were located in the left ($x = -54$, $y = -26$, $z = +4$) and right ($x = +52$, $y = -16$, $z = +2$) STG as revealed in the distributed source reconstruction of the matching < mismatching effect and had a SD in each direction of 5 mm. Across all gradiometer sensors, the group mean percentage of variance explained by the resulting ECD models was 75 ± 11% for the effect of prior knowledge and 77 ± 9% for the effect of perceptual learning. This analysis also was run with looser constraints on the prior locations (SDs of 7.07 and 10 mm), but there was no significant interaction between effect type (prior knowledge/perceptual learning) and degree of constraint (5-/7-/10-mm SD) on ECD locations (all *P*s > 0.1). Furthermore, when ANOVAs of dipole locations were conducted separately for each degree of constraint, the same main effects (of effect type and hemisphere) and interactions were significant across all degrees of constraint. Thus, reported results (using SD = 5 mm) are robust over a wide range of constraints on ECD locations. Source activity for the three critical contrasts (sensory detail, prior knowledge, and perceptual learning) was extracted from dipoles based on the mean location and orientation across ECD models and participants by using the inverse of each participant's forward model (which maps activity from sources to sensors).

ACKNOWLEDGMENTS. We thank Maarten van Casteren, Clare Cook, and Lucy MacGregor for excellent technical support during data collection; Nicolas Barascud for providing Matlab code to plot source dipole locations; and Helen Blank, Thomas Cope, two anonymous reviewers, and the editor for helpful comments on an earlier draft of this manuscript. This research was supported by a Medical Research Council (MRC) Centenary Award (to E.S.) and MRC Grant MC-A060-5PQ80 (to M.H.D.).

1. Remez RE, Rubin PE, Pisoni DB, Carrell TD (1981) Speech perception without traditional speech cues. *Science* 212(4497):947–949.
2. Jacoby LL, Allan LG, Collins JC, Larwill LK (1988) Memory influences subjective experience: Noise judgments. *J Exp Psychol Learn Mem Cogn* 14:240–247.
3. Ludmer R, Dudai Y, Rubin N (2011) Uncovering camouflage: Amygdala activation predicts long-term memory of induced perceptual insight. *Neuron* 69(5):1002–1014.
4. Dolan RJ, et al. (1997) How the brain learns to see objects and faces in an impoverished context. *Nature* 389(6651):596–599.
5. Goldstone RL (1998) Perceptual learning. *Annu Rev Psychol* 49:585–612.
6. Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen* 134(2):222–241.
7. Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8(10):457–464.
8. Ahissar M, Hochstein S (1997) Task difficulty and the specificity of perceptual learning. *Nature* 387(6631):401–406.
9. McQueen JM, Cutler A, Norris D (2006) Phonological abstraction in the mental lexicon. *Cogn Sci* 30(6):1113–1126.
10. Fenn KM, Nusbaum HC, Margoliash D (2003) Consolidation during sleep of perceptual learning of spoken language. *Nature* 425(6958):614–616.
11. Polat U, Ma-Naim T, Belkin M, Sagi D (2004) Improving vision in adult amblyopia by perceptual learning. *Proc Natl Acad Sci USA* 101(17):6692–6697.
12. Moore DR, Shannon RV (2009) Beyond cochlear implants: Awakening the deafened brain. *Nat Neurosci* 12(6):686–691.

13. Stacey PC, et al. (2010) Effectiveness of computer-based auditory training for adult users of cochlear implants. *Int J Audiol* 49(5):347–356.
14. Norris D, McQueen JM, Cutler A (2003) Perceptual learning in speech. *Cognit Psychol* 47(2):204–238.
15. Rubin N, Nakayama K, Shapley R (1997) Abrupt learning and retinal size specificity in illusory-contour perception. *Curr Biol* 7(7):461–467.
16. Mitterer H, McQueen JM (2009) Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One* 4(11):e7785.
17. Jesse A, McQueen JM (2011) Positional effects in the lexical retuning of speech perception. *Psychon Bull Rev* 18(5):943–950.
18. Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP (2008) Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *J Exp Psychol Hum Percept Perform* 34(2):460–474.
19. Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360(1456):815–836.
20. Mirman D, McClelland JL, Holt LL (2006) An interactive Hebbian account of lexically guided tuning of speech perception. *Psychon Bull Rev* 13(6):958–965.
21. Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK (2010) Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J Neurosci* 30(21):7179–7186.
22. Erb J, Henry MJ, Eisner F, Obleser J (2013) The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J Neurosci* 33(26):10688–10697.
23. Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. *J Neurosci* 32(25):8443–8453.
24. Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang Cogn Process* 27:1145–1166.
25. Myers EB, Mesite LM (2014) Neural systems underlying perceptual adjustment to non-standard speech tokens. *J Mem Lang* 76:80–93.
26. Kilian-Hütten N, Vroomen J, Formisano E (2011) Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *Neuroimage* 57(4): 1601–1607.
27. Howes D (1957) On the Relation between the Intelligibility and Frequency of Occurrence of English Words. *J Acoust Soc Am* 29:296–305.
28. Pollack I (1959) Intelligibility of Known and Unknown Message Sets. *J Acoust Soc Am* 31:273–279.
29. Miller GA, Isard S (1963) Some perceptual consequences of linguistic rules. *J Verbal Learn Verbal Behav* 2:217–228.
30. Rosen S, Faulkner A, Wilkinson L (1999) Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J Acoust Soc Am* 106(6): 3629–3636.
31. Clarke CM, Garrett MF (2004) Rapid adaptation to foreign-accented English. *J Acoust Soc Am* 116(6):3647–3658.
32. Peelle JE, Wingfield A (2005) Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J Exp Psychol Hum Percept Perform* 31(6):1315–1330.
33. Huyck JJ, Johnsrude IS (2012) Rapid perceptual learning of noise-vocoded speech requires attention. *J Acoust Soc Am* 131(3):EL236–EL242.
34. Dorjee D, Bowers JS (2012) What can fMRI tell us about the locus of learning? *Cortex* 48(4):509–514.
35. Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2014) Top-down influences of written text on perceived clarity of degraded speech. *J Exp Psychol Hum Percept Perform* 40(1): 186–199.
36. Friston KJ, Henson RN (2006) Commentary on: Divide and conquer; a defence of functional localisers. *Neuroimage* 30:1097–1099.
37. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12(5):535–540.
38. Kilner JM (2013) Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clin Neurophysiol* 124(10):2062–2063.
39. Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23(8):3423–3431.
40. Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* 28(32):8116–8123.
41. Scott SK, Rosen S, Lang H, Wise RJS (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech–a positron emission tomography study. *J Acoust Soc Am* 120(2):1075–1083.
42. Kiebel SJ, Daunizeau J, Phillips C, Friston KJ (2008) Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. *Neuroimage* 39(2):728–741.
43. Eickhoff SB, et al. (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25(4):1325–1335.
44. Yvert B, Fischer C, Bertrand O, Pernier J (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage* 28(1):140–153.
45. Sumby WH (1954) Visual Contribution to Speech Intelligibility in Noise. *J Acoust Soc Am* 26(2):212–215.
46. Frauenfelder UH, Segui J, Dijkstra T (1990) Lexical effects in phonemic processing: Facilitatory or inhibitory. *J Exp Psychol Hum Percept Perform* 16(1):77–91.
47. Wild CJ, Davis MH, Johnsrude IS (2012) Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage* 60(2):1490–1502.
48. Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12(6):718–724.
49. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.
50. Alain C, Campeanu S, Tremblay K (2010) Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J Cogn Neurosci* 22(2):392–403.
51. Ross B, Tremblay K (2009) Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hear Res* 248(1-2):48–59.
52. Tremblay KL, Ross B, Inoue K, McClannahan K, Collet G (2014) Is the auditory evoked P2 response a biomarker of learning? *Front Syst Neurosci* 8:28.
53. Mitterer H, Reinisch E (2013) No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *J Mem Lang* 69(4):527–545.
54. Grossberg S (1987) Competitive learning: From interactive activation to adaptive resonance. *Cogn Sci* 63:23–63.
55. Bastos AM, et al. (2012) Canonical microcircuits for predictive coding. *Neuron* 76(4): 695–711.
56. Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16(7):390–398.
57. Spratling MW (2008) Reconciling predictive coding and biased competition models of cortical function. *Front Comput Neurosci* 2:4.
58. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36(3):181–204.
59. Henson RN, Gagnepain P (2010) Predictive, interactive multiple memory systems. *Hippocampus* 20(11):1315–1326.
60. Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22(7):615–621.
61. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87.
62. Moran RJ, et al. (2013) Free energy, precision and learning: The role of cholinergic neuromodulation. *J Neurosci* 33(19):8227–8236.
63. Petrov AA, Dosher BA, Lu Z-L (2005) The dynamics of perceptual learning: An incremental reweighting model. *Psychol Rev* 112(4):715–743.
64. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270(5234):303–304.
65. Baayen RH, Piepenbrock R, Van Rijn H (1993) *The CELEX Lexical Database* (CD-ROM) (University of Pennsylvania, Linguistic Data Consortium, Philadelphia).
66. Levenshtein V (1966) Binary codes capable of correcting deletions, insertions and reversals. *Sov Phys Dokl* 10(8):707–710.
67. Taulu S, Simola J, Kajola M (2005) Applications of the signal space separation method. *IEEE Trans Signal Process* 53:3359–3372.
68. Wager TD, Keller MC, Lacey SC, Jonides J (2005) Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage* 26(1):99–113.
69. Litvak V, et al. (2011) EEG and MEG data analysis in SPM8. *Comput Intell Neurosci* 2011:852961.
70. Kilner JM, Friston KJ (2010) Topological inference for EEG and MEG. *Ann Appl Stat* 4:1272–1290.
71. Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE (2004) Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage* 22(2): 676–687.
72. Henson RN, Mouchlianitis E, Friston KJ (2009) MEG and EEG data fusion: Simultaneous localisation of face-evoked responses. *Neuroimage* 47(2):581–589.
73. Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ (2005) An empirical Bayesian solution to the source reconstruction problem in EEG. *Neuroimage* 24(4):997–1011.
74. Litvak V, Friston K (2008) Electromagnetic source reconstruction for group studies. *Neuroimage* 42(4):1490–1498.
75. Henson RN, Wakeman DG, Litvak V, Friston KJ (2011) A parametric empirical Bayesian framework for the EEG/MEG inverse problem: Generative models for multi-subject and multi-modal integration. *Front Hum Neurosci* 5:76.
76. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation. *Neuroimage* 34(1):220–234.
77. Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int J Psychophysiol* 18(1):49–65.
78. Friston K, Henson R, Phillips C, Mattout J (2006) Bayesian estimation of evoked and induced responses. *Hum Brain Mapp* 27(9):722–735.
79. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66(3):241–251.
80. Hansen P, Kringelbach M, Salmelin R, eds (2010) *MEG: An Introduction to Methods* (Oxford Univ Press, New York).
81. Fiser J, Berkes P, Orbán G, Lengyel M (2010) Statistically optimal perception and learning: From behavior to neural representations. *Trends Cogn Sci* 14(3):119–130.
82. Knill DC, Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci* 27(12):712–719.
83. McGettigan C, Rosen S, Scott SK (2014) Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. *Front Syst Neurosci* 8:18.
84. Azadpour M, Balaban E (2015) A proposed mechanism for rapid adaptation to spectrally distorted speech. *J Acoust Soc Am* 138(1):44–57.
85. Loftus GR, Masson MEJ (1994) Using confidence intervals in within-subject designs. *Psychon Bull Rev* 1(4):476–490.

Supporting Information

Sohoglu and Davis 10.1073/pnas.1523266113

SI Methods

In this predictive coding account of the perception of degraded speech (depicted in Fig. 6*A*), perceptual and neural outcomes are determined by interactions between two hierarchically organized levels of representation: sensory (acoustic–phonetic) features (assumed to be represented in the STG) and categorical phonological representations (in higher-level inferior frontal and precentral gyri). Simulations of these interactions were informed by standard predictive coding views of perception (19, 55–61) in which perceptual hypotheses generate predictions for expected sensory input, and prediction errors (the difference between expected and actual sensory input) are used to update perceptual hypotheses. These updated perceptual hypotheses (Fig. 6*B*) are assumed to correlate with behavioral outcomes (clarity ratings or recognition accuracy). The summed absolute magnitude of the prediction error (shown in Fig. 6*C*) is assumed to correlate with the magnitude of neural responses measured in the STG in our experiment. Our simulation therefore follows other predictive coding accounts (19, 55, 60, 79) in assuming that prediction error signals within neocortical hierarchies are generated by large pyramidal neurons found in superficial cortical laminae. These neurons are proposed to be a significant contributor to the MEG signal because their dendrites are aligned and oriented perpendicular to the cortical surface (80).

In our simulation of degraded speech perception, we explored how a single sensory feature (e.g., VOT) could be used to distinguish between two phonological categories (e.g., voiced and unvoiced segments such as /b/ and /p/). Feature and phonological levels of representation were both modeled by assigning activation values to a set of units that represent a probability density function (PDF) as depicted in the bar graphs of Fig. 6*A* and Fig. S4). These PDFs might be instantiated neurally as population codes (81, 82). For example, each unit along the *x* axis of a PDF could represent a cortical region maximally responsive to the phonological category or sensory feature indicated by its position on the *x* axis. Thus, hearing the segment /b/ on a single trial would produce maximal activation in a cortical region tuned to that segment (corresponding to the peak of the PDF) and also would produce activation in other regions tuned to perceptually similar representations (corresponding to the tails of the PDF). There were two units representing each of the two phonological categories (A and B), whereas sensory feature representations were more graded or continuous, encoding the distribution of likely feature values over 21 arbitrarily scaled units.

Despite the considerable oversimplification in modeling the processing of only a single sensory feature, VOT, our simulation illustrates how predictive coding computations can explain the impact of three key experimental manipulations of behavioral and neural responses: changes in sensory detail, (mis)matching prior knowledge, and perceptual learning. All three factors can enhance perceptual outcomes in a similar way (shown by the updated state of perceptual hypotheses after processing new sensory input) but differentially impact neural responses (i.e., the summed absolute magnitude of prediction error).

Distributions of sensory features in the input were generated from two underlying PDFs, one for each phonological category, with the amount of sensory detail determining the area under the two curves (i.e., the probability of each category being present in the input). For highly degraded (one-channel) speech, the two categories had nearly equally probabilities (0.55 vs. 0.45). For clear (24-channel) speech, the two categories had very different probabilities (0.95 vs. 0.05). For six-channel speech, the level of sensory detail used to simulate manipulations of prior knowledge and perceptual learning (see below), intermediate values (0.75 vs. 0.25) were specified. These parameter values are also listed in Table S1.

When simulating the effects of sensory detail (24 channels versus one channel) and perceptual learning (post- versus pretest phases), predictions for the current sensory input were neutral (i.e., both phonological categories were equally likely, with a probability of 0.5, reflecting the absence of strong prior knowledge from written text in these conditions). However, when simulating effects of prior knowledge (matching vs. mismatching), predictions were biased toward one or other category (i.e., having a probability of 0.75 for the predicted category and 0.25 for the nonpredicted category; chosen to reflect the likelihood of prior expectations from written text matching the sensory signal in the experiment). These perceptual hypotheses were multiplied by a 2 × 21 element weight matrix to generate a distribution of predicted sensory features associated with these perceptual hypotheses. These weights specify two PDFs that expressed the mean value of the sensory feature and the SD or precision of the predicted sensory feature associated with each phonological category (see Table S1 for parameter values used to generate these weights).

In simulating perceptual learning, reductions in prediction error were attributed to changes in the variance or precision of predictions for sensory features. Optimal perceptual outcomes and minimal prediction errors occurred when the precision of sensory predictions matched the precision of the sensory input (62). We therefore simulated perceptual learning by contrasting perceptual outcomes and prediction errors, during a pretraining period in which the distribution of sensory features was more precise than predicted, with a posttraining period in which predictions were made with an increased precision that matched the sensory input (i.e., we used identical parameters for the SD of the category-to-feature weights and the sensory input in Table S1). This change in precision had the effect of increasing the amount of information in updated perceptual hypotheses and hence the accuracy of perceptual outcomes while still decreasing the magnitude of prediction error and hence the magnitude of the STG response. These changes therefore are in line with the behavioral and neural observations in our experiment.

SI Discussion

In functional terms, an increase in the precision of sensory predictions has the effect of increasing the amount of information gained from sensory features. This increase in precision therefore provides a neural implementation of "attentional weighting" theories (5) in which perceptual learning derives from using informative features during perception more appropriately while down-weighting uninformative features (7, 62, 63). We note that previous studies of perceptual learning of vocoded speech have shown that enhanced consonant identification is associated with increased information transmission for voicing and manner but not for place of articulation features (83). This finding is consistent with an attentional weighting account because it also has been shown that place features are more degraded by vocoding than voicing or manner features (64). Similar changes in information transmission—upweighting of informative phonetic features—have been shown for perceptual learning of spectrally rotated speech (84).

A significant contribution of the present experimental work is that we show a cross-subject correlation between the magnitude of the reduction of neural activity by prior knowledge during training and the magnitude of perceptual learning (improved speech recognition accuracy) for post- versus pretraining test sessions. In our simulation, trials in which degraded speech follows matching text are associated with both a distinctive behavioral outcome (i.e., confident perceptual identification) and a triphasic profile of sensory prediction errors (Fig. S4*B*, matching trials). A mechanism by which our simulation could explain the observed cross-subject correlation is that, after successful perception, a Hebbian learning rule modifies the weights that link phonological representations and sensory features. For this account, the direction and magnitude of weight changes is determined by the magnitude and sign of sensory prediction error. The triphasic pattern of prediction error (Fig. S4*B*, matching trials) therefore would result in weight changes that lead to a narrower distribution of predicted sensory features on subsequent trials (i.e., increased precision for sensory predictions)—exactly the weight changes that differentiate the pretraining and posttraining test sessions.

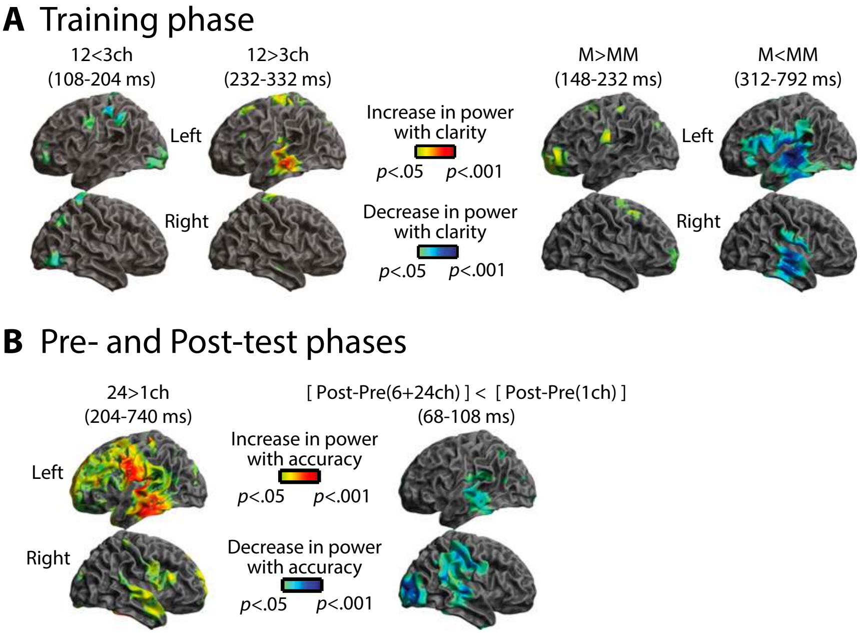

Fig. S1. Distributed source reconstruction of sensor-space effects overlaid onto a template brain (using data from all sensors, including EEG and MEG magnetometers and gradiometers). The upper and lower rows show lateral views of the left and right hemispheres, respectively. (*A*) Training phase. Red and blue colors indicate increases and decreases in source power for enhanced speech clarity ratings, respectively. (*B*) Pre- and posttest phases. Red and blue colors indicate increases and decreases in source power for enhanced speech recognition accuracy, respectively. ch, channels; M, matching; MM, mismatching.

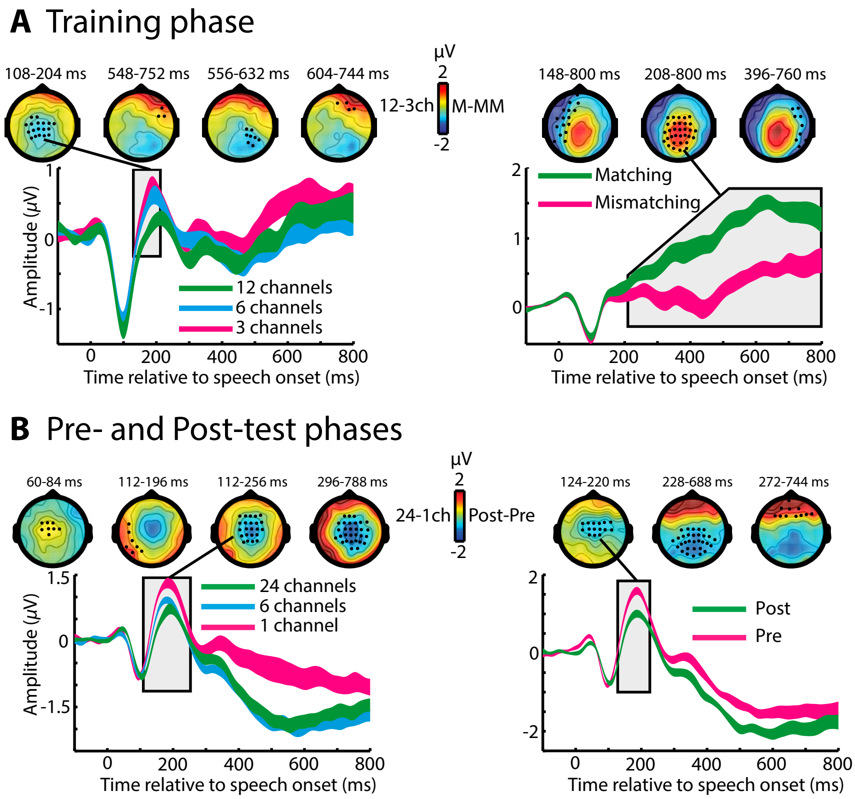

Fig. S2. Group-level effects in EEG sensors, plotted as in Fig. 3. (*A*) Training phase. (*B*) Pre- and posttest phases. ch, channels; M, matching; MM, mismatching.

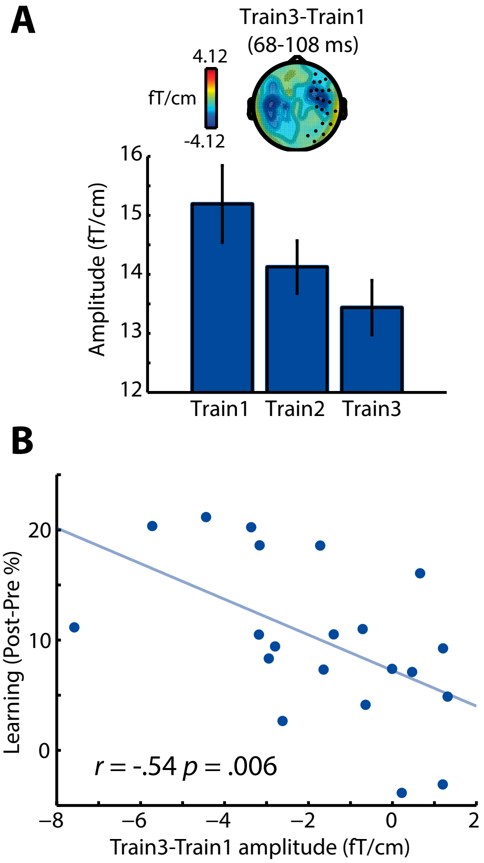

Fig. S3. (*A*) Effects of block number on neural responses during the training phase. From 68–108 ms after speech onset, the MEG response is reduced in magnitude across the three blocks of the training phase (similar to the post- vs. pretest effect) in sensors shown as black circles on the topographic plot. Error bars indicate ± two within-subject SEMs; the topography represents the difference in MEG response between training blocks 3 and 1. (*B*) The difference in MEG (gradiometer) response between training blocks 3 and 1 also correlated across subjects with the magnitude of learning-related change in behavior (improvement in speech recognition post- vs. pretest for six-channel speech).

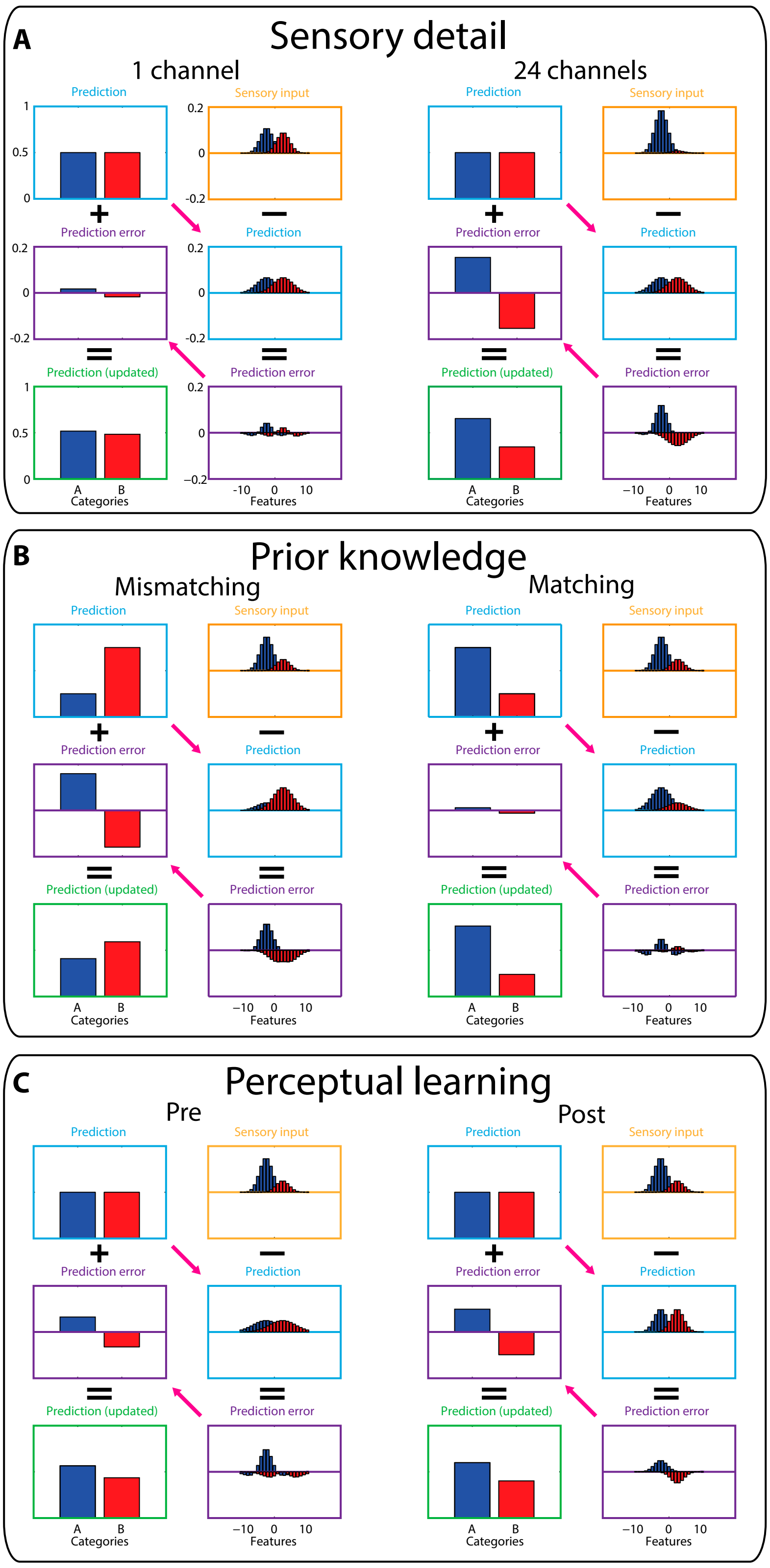

Fig. S4. Predictive coding simulations of key experimental conditions. For an overview of the simulation method, see Fig. 6 and its accompanying legend and *SI Methods*. A full set of parameters used for these simulations is listed in Table S1. (*A*) Sensory detail simulations (24 channels versus one channel during pre- and posttest phases). In these conditions listeners heard speech without prior written text, and hence phonological categories A and B were equally predicted (light blue boxes, equivalent to a neutral prior). As sensory detail increased from one channel to 24 channels, the sensory input more clearly favored category A (orange boxes), leading to increased feature prediction error (shown in purple boxes, linked to neural responses in the STG) and, in turn, to updated predictions and more accurate perceptual outcomes (shown in green boxes, linked to clarity or recognition accuracy measures). (*B*) Prior knowledge simulations (matching vs. mismatching conditions, six-channel speech heard during the training phase). In these simulations speech provided intermediate sensory evidence for category A (orange boxes) which was inconsistent (mismatching trials) or consistent (matching trials) with prior predictions (light blue boxes). Mismatching prior predictions resulted in a large discrepancy between predicted and actual sensory input and hence large errors in feature prediction (purple boxes) associated with larger STG responses. In the mismatching condition, these prediction errors were insufficient to overcome the strong prediction for category B leading to inaccurate perceptual outcomes (green box). By comparison, in the matching condition there was good correspondence between predicted and actual sensory input (because both favored category A), resulting in small errors in feature prediction (purple box) and more accurate perceptual outcomes (green box). (*C*) Simulated effects of perceptual learning (post- versus pretest phases, six-channel speech). As in the manipulation of prior knowledge, errors in feature prediction were reduced in post- vs. pretest phases because of better correspondence between predicted and actual sensory input. However, this reduction in prediction error arose even though speech in the pre- and posttest conditions was presented without prior written text and therefore without strong predictions for specific categories, as in panel *A*. Rather, perceptual learning arose from changes in connection weights that map between categories and features (shown as pink arrows), which increase their precision (i.e., reduce their variance) to match better the precision (variance) of the sensory input. Before training, predictions for sensory features (light blue box) came from a more variable distribution than the sensory input and therefore generated nonoptimal prediction errors (purple box). In effect, these predictions underestimated the informativeness of the sensory input, and therefore updated predictions (green box) were not fully accurate. Posttraining, however, the distribution of predicted feature values (light blue box) more closely approximated the sensory input (orange box), reflecting a more accurate estimate of the informativeness of the sensory input. These more accurate predictions resulted in reduced feature prediction error (purple box) and more accurate perceptual outcomes (green box). Thus, perceptual learning in this simulation arises through more appropriate weighting of sensory features in degraded speech signals (see *SI Methods* for details).

Table S1. Full parameters for computational simulations described in Fig. 6 and Fig. S4 and associated legends

	Condition					
Simulation parameters	1 ch	24 ch	MM (6 ch)	M (6 ch)	Pretest (6 ch)	Posttest (6 ch)
Prediction (categories)						
P A (B)	0.5 (0.5)	0.5 (0.5)	0.25 (0.75)	0.75 (0.25)	0.5 (0.5)	0.5 (0.5)
Category-to-feature weights						
Mean A (B)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)
SD A (B)	3 (3)	3 (3)	3 (3)	3 (3)	4 (4)	2 (2)
Sensory input (features)						
P A (B)	0.55 (0.45)	0.95 (0.05)	0.75 (0.25)	0.75 (0.25)	0.75 (0.25)	0.75 (0.25)
Mean A (B)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)	−2.5 (+2.5)
SD A (B)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)
Other parameters						
Update: 0.5						
Feature set: −10:+10						

See *SI Methods* for additional details. The first numerical values in each cell concern category A, and those in parentheses concern category B. *P* A (B) specifies the relative probability (activation) for category A (or B) at the phonological level. The category-to-feature weights that converted these phonological representations into sensory feature representations at a lower level (and vice-versa, that converted sensory prediction errors into phonological prediction errors) are specified as PDFs, each with a mean and SD. Activations in sensory input units also are specified as PDFs (one for each of the two categories) characterized by mean and SD parameters. The relative area under each of the two PDFs represents the probability of each category being present in the input. These relative areas were used to simulate the manipulation of sensory detail, with increases in sensory detail leading to stronger sensory evidence for one or the other category. The update parameter served to scale how much phonological predictions were updated in response to phonological prediction error. The feature set parameter is a vector that specifies the range of units over which the PDFs were computed (essentially, the range of values in the *x* axis on the feature graphs in Fig. 6 and Fig. S4). ch, channels; M, matching; MM, mismatching.