Behavioral/Cognitive

The Brain Dynamics of Rapid Perceptual Adaptation to Adverse Listening Conditions

Julia Erb,[1] **Molly J. Henry,**[1] **Frank Eisner,**[2] **and Jonas Obleser**[1]

[1]Max Planck Research Group "Auditory Cognition," Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany and [2]Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands

Listeners show a remarkable ability to quickly adjust to degraded speech input. Here, we aimed to identify the neural mechanisms of such short-term perceptual adaptation. In a sparse-sampling, cardiac-gated functional magnetic resonance imaging (fMRI) acquisition, human listeners heard and repeated back 4-band-vocoded sentences (in which the temporal envelope of the acoustic signal is preserved, while spectral information is highly degraded). Clear-speech trials were included as baseline. An additional fMRI experiment on amplitude modulation rate discrimination quantified the convergence of neural mechanisms that subserve coping with challenging listening conditions for speech and non-speech. First, the degraded speech task revealed an "executive" network (comprising the anterior insula and anterior cingulate cortex), parts of which were also activated in the non-speech discrimination task. Second, trial-by-trial fluctuations in successful comprehension of degraded speech drove hemodynamic signal change in classic "language" areas (bilateral temporal cortices). Third, as listeners perceptually adapted to degraded speech, downregulation in a cortico-striato-thalamo-cortical circuit was observable. The present data highlight differential upregulation and downregulation in auditory–language and executive networks, respectively, with important subcortical contributions when successfully adapting to a challenging listening situation.

Introduction

Humans have the capability to rapidly adapt to degraded or altered speech. This challenge is particularly relevant to cochlear implant (CI) patients adapting to an extremely distorted auditory input delivered by their device (Giraud et al., 2001; Fallon et al., 2008). The current study investigated the short-term neural processes that underlie this adaptation to degraded speech using functional magnetic resonance imaging (fMRI).

We simulated CI-transduced speech in normal-hearing listeners using noise-vocoding (Shannon et al., 1995), which degrades the spectral detail in an auditory signal, and forces the listener to rely more on the (intact) temporal envelope cues for speech comprehension. Listeners with higher sensitivity to envelope fluctuations in an auditory signal, as measured by amplitude modulation (AM) rate discrimination thresholds, adapt more quickly to vocoded speech (Erb et al., 2012). Therefore, we predicted that temporal envelope of non-speech sounds (Giraud et al., 2000) and vocoded speech should be processed by shared neural resources.

Rapid perceptual learning of vocoded speech is a well established finding of behavioral studies (Rosen et al., 1999; Davis et al., 2005; Peelle and Wingfield, 2005; Bent et al., 2009), but its neural bases are largely unknown. Recently, increased precentral gyrus activity was associated with boosted perceptual learning during the joint presentation of vocoded and clear words (Hervais-Adelman et al., 2012). Further, in a simulated CI rehabilitation program that supplemented presentation of vocoded sentences with simultaneous written feedback, perceptual learning relied on the left inferior frontal gyrus (IFG; Eisner et al., 2010). However, in everyday situations, listeners rarely receive direct feedback. Therefore, the neural dynamics of feedback-free or self-regulated adaptation are investigated here.

Subcortical structures likely play a critical role in adaptation to degraded speech, but the specific contributions of distinct structures are unclear. A recent voxel-based morphometry study demonstrated that gray matter volume in the left pulvinar thalamus predicted how fast listeners adapted to vocoded speech (Erb et al., 2012). However, previous fMRI studies have failed to detect adaptation-related signal changes in subcortical regions, possibly due to these brain areas' susceptibility to MR artifacts. Here, we implemented a cardiac-gated image acquisition procedure, thereby avoiding heartbeat-related artifacts.

The present fMRI study sheds new light on perceptual adaptation to degraded speech with respect to four points: (1) we investigate the convergence of neural mechanisms underlying effortful speech and non-speech processing; (2) we test feedback-free, short-term adaptation; (3) unlike previous perceptual learning studies (Golestani and Zatorre, 2004; Adank and Devlin, 2010; Eisner et al., 2010), we collect word report scores on every trial and are able to model both trial-by-trial fluctuation in comprehension and adaptation to degraded speech; and (4) we assess subcortical contributions to adaptation. We show that an "exec-
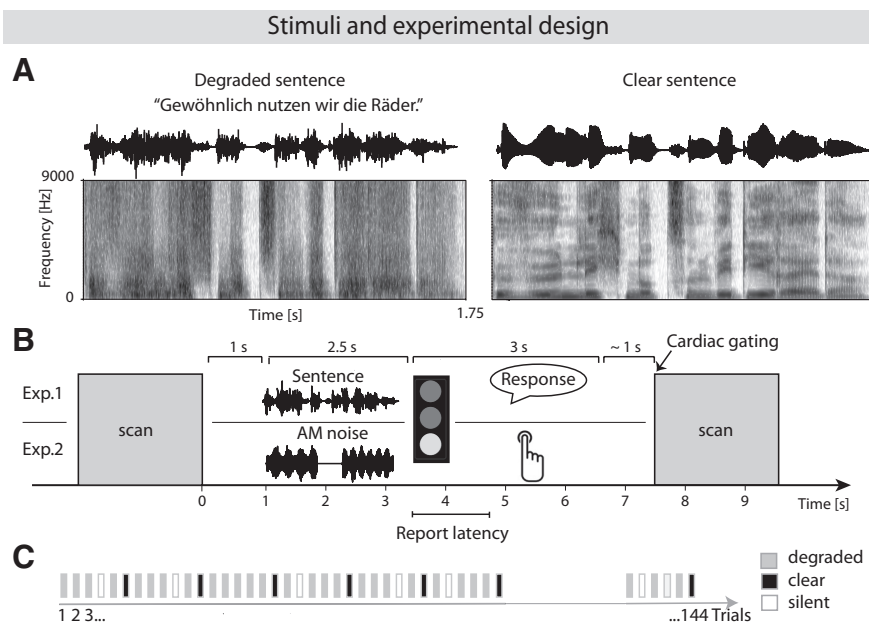
Figure 1. Stimuli and experimental design. *A*, Oscillogram (top) and spectrogram (bottom) of the sentence "Gewöhnlich nutzen wir die Räder" (Normally we use the bikes) in degraded (left) and clear speech (right). *B*, Trial structure in Experiment 1 and 2. After acoustic presentation of a sentence or AM stimulus (lasting for ~2.5 s), participants were prompted to respond when green lights appeared on a screen (lasting for 3 s) by repeating the sentence or pressing a button. Subsequent scan acquisition was initiated by cardiac gating. The onset of auditory stimulation preceded the anticipated scan acquisition by a period of 6.5 s, although the actual scan time was variable due to cardiac gating. *C*, Experiment 1 comprised 100 degraded speech trials, 24 clear speech trials (inserted after every fifth trial), and 20 randomly interspersed silent trials.

Butterworth lowpass filter with a cutoff frequency of 400 Hz. Noise-vocoding was applied to all sentences in MATLAB 7.11 as described in Rosen et al. (1999) using four frequency bands spanning 70–9000 Hz that were spaced according to Greenwood's cochlear frequency-position function (Greenwood, 1990; for full settings see Erb et al., 2012). The waveform and spectrogram of the vocoded and clear version of an example sentence are shown in Figure 1A.

On each trial, participants first heard a sentence. They were instructed to repeat as much of the sentence as they had understood when a green light appeared on the screen, but to stop talking when the green light disappeared (after 3 s) to avoid movement during scan acquisition (Fig. 1B). Speech production was recorded for later off-line scoring (Eckert et al., 2009; Harris et al., 2009). Responses were scored as proportions of correctly repeated words per sentence ("report scores"; Peelle et al., 2013). Scoring took into account all words of a sentence including function words; errors in declension or conjugation were accepted as correct.

Clear speech trials were used as a high-level baseline. Clear speech can be assumed to be fully adapted, and therefore to be processed in a stable way over time. This ensured that no neural adaptation would occur in the baseline condition, whereas another type of artificial speech degradation (e.g., rotated speech) might have led to neural adaptation (even in the absence of behavioral adaptation).

In sum, Experiment 1 comprised three conditions: (1) 4-band vocoded sentences ("degraded speech"; 100 trials in total), (2) clear (non-vocoded) sentences ("clear speech"; 24 trials in total), and (3) trials lacking any auditory stimulation ("silent trials"; 20 trials in total). Overall, the experiment comprised 144 trials. Clear speech trials were presented every fifth trial, whereas the silent trials were randomly interspersed (Fig. 1C). Sentences were presented to each participant in one of four different orders; presentation order was counterbalanced across participants.

Participants' adaptation curves. As in Erb et al. (2012), we modeled each individual's performance improvement in two different ways: as a power law and as a linear performance increase. To test which function would better describe the data, both curves were fitted to the individual report scores over time using a least-squares estimation procedure in MATLAB 7.11 (for example fits to the scores averaged over participants, see Fig. 2A). We compared goodness of fit by determining the Bayesian information criterion (BIC; Schwarz, 1978) of the linear and the power law fits within each participant.

Experiment 2. Stimuli were 1 s long sinusoidally amplitude-modulated white noises. The standard stimulus was modulated at 4 Hz. Deviant stimuli were modulated at seven different rates that were linearly spaced between 2 and 6 Hz in steps of 0.67 Hz. The middle level was modulated at the same rate as the standard (4 Hz); critically, in this condition participants were unable to distinguish standard and deviant, but still performed the same task. Modulation depth was kept constant at 60%. The onset phase of the sinusoidal modulation was randomly varied for all stimuli, separately for the standard and deviant. Standard and deviant stimuli were presented with an interstimulus interval of 500 ms.

The paradigm was a two-alternative forced choice task: on each trial, participants first heard the standard stimulus (modulated at 4 Hz) followed by one deviant stimulus. After auditory presentation of the sound pair, participants were prompted to respond when a green light appeared on the screen (Fig. 1B). Half of the participants were to indicate which sound had a faster modulation rate, while the other half's task was to

utive" system (Eckert et al., 2009) is recruited when actively coping with difficult speech and non-speech listening conditions. In contrast, activations in the classic "language" network (Scott et al., 2000; Hickok and Poeppel, 2007) are driven by trial-by-trial fluctuations in speech comprehension, not acoustic speech clarity. Finally, we demonstrate that rapid adaptation to degraded speech is accompanied by hemodynamic downregulation in a cortico-striato-thalamo-cortical network.

Materials and Methods

Participants
Thirty participants (15 females, age range 21–31 years, mean 25.9 years) took part in the study. Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences. All were native speakers of German with no known hearing impairments, language or neurological disorders and showed dominant right-handedness according to the Edinburgh inventory (Oldfield, 1971). They were naive to noise-vocoded speech. Participants received financial compensation of €16, and gave informed consent. Procedures were approved by the local ethics committee (University of Leipzig).

Stimuli and experimental design
Experiment 1. Sentence material was drawn from a German version of the speech in noise (SPIN) sentences (Kalikow et al., 1977; Erb et al., 2012), which is controlled for the predictability of the final word (high vs low predictability). For the present study, only low-predictability sentences were chosen, such that semantic cues were limited and the listener had to rely primarily on acoustic properties of the sentence for comprehension. A complete list of these sentences is available in Erb et al. (2012).

The sentences were recorded by a female speaker of standard German in a sound proof chamber. The length of the recorded sentences varied between 1620 and 2760 ms. Sentences were degraded using 4-band noise vocoding. This procedure divides the raw signal into frequency bands, extracts the amplitude envelope from each band, and reapplies it to bandpass-filtered noise carriers, thereby smearing spectral fine structure. For envelope extraction, we used a second-order, zero-phase

indicate which sound had a slower modulation rate to counterbalance button presses. Participants responded via a button box in their right hand by pressing the left (for first sound) or right key (for second sound).

In sum, there were seven levels of deviant AM rate (comprising 16 trials each); in addition, we interspersed silent trials lacking any auditory stimulation (16 trials). On the whole, every participant listened to 128 trials in a pseudorandom order where trials of the same condition were never presented subsequently.

Experimental procedure

To maximize the comparability between individuals, all participants were tested in the same order, namely degraded speech perception first (Experiment 1), followed by AM rate discrimination (Experiment 2). Before participants went into the scanner, they were familiarized with each of the two tasks; they listened to three 8-band vocoded GSPIN sentences as training for Experiment 1 and three exemplary trials of Experiment 2.

In the scanner, to prevent hearing damage due to scanner noise, participants wore Alpine Musicsafe Pro earplugs, yielding approximately linear 14 dB reduction in sound pressure up to 8 kHz. Auditory stimuli were delivered through MR-Confon headphones using Presentation software. Visual prompts were projected on a screen which participants viewed via a mirror.

Trial timing was identical in Experiment 1 and 2. Each trial was ~9 s long, but actual trial length varied due to cardiac gating (see below). Trials started with a 1 s silent gap, after which participants heard a sentence (Experiment 1) or two AM-modulated stimuli (Experiment 2) lasting for ~2.5 s. Following stimulus presentation (3.5 s into the trial), a green light ("go"-signal for response) was visually presented and lasted for 3 s. After ~1 s of silence, scan acquisition with a TR of 2 s was triggered using cardiac gating. Thus, the onset of auditory stimulation preceded the anticipated scan acquisition by ~6.5 s (Fig. 1B).

MRI data acquisition

MRI data were collected on a 3 T Siemens Verio scanner. Blood oxygenation level-dependent (BOLD) fMRI images were acquired with a 12-channel head coil using an echo-planar imaging (EPI) sequence [TR ≈ 9000 ms, TA = 2000 ms, TE = 30 ms, flip angle = 90°, 3 mm slice thickness, 30 axial slices (ascending), interslice distance = 1 mm, acquisition matrix of 64 × 64, voxel size = 3 × 3 × 3 mm] in two separate runs for Experiment 1 and 2. The acquisition matrix was placed such that the x-axis was in line with the anterior–posterior commissure (AC–PC). We used a sparse-sampling procedure, where TR was longer than TA, allowing for silent periods to play stimuli and record verbal responses (Hall et al., 1999).

Additionally, cardiac gating was applied to avoid movement artifacts caused by the heartbeat in subcortical structures (von Kriegstein et al., 2008), in which we were especially interested. Participants' heartbeat was monitored using an MR-compatible pulse oximeter (Siemens) attached
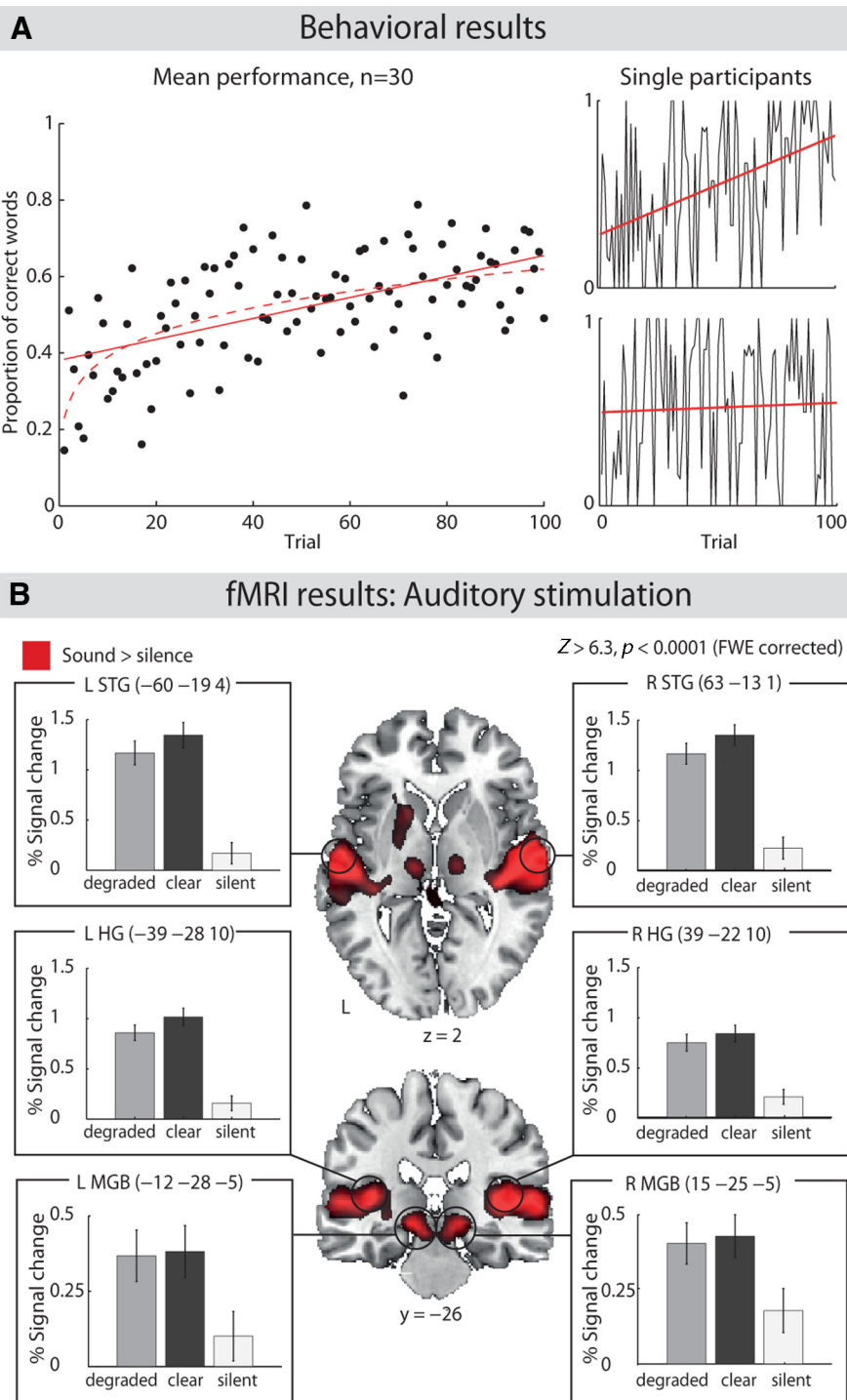
to their left ring finger. On each trial, after 9 s had elapsed, the scanner waited for the first heartbeat to trigger volume acquisition. Thus, the actual repetition time (TR) was variable but amounted to 9.45 ± 0.27 s (mean ± SEM; across all participants).

Following functional imaging, a T1-weighted structural image was acquired with a 32-channel head coil using an MPRAGE sequence [TR = 1300 ms, TE = 3.5 ms, flip angle = 10°, 1 mm slice thickness,

Figure 2. ***A***, Behavioral adaptation to degraded speech. Linear (red solid line) and power law fit (dashed line) to mean report scores averaged over participants (left). Two examples of participants with a steep slope (top) and a shallow slope (bottom) in their linear fits ("adaptation curves"). Although there is a general increase in performance, it is clearly visible that report scores strongly fluctuate from trial to trial (right). ***B***, Effects of auditory stimulation in Experiment 1. Auditory cortices (HG, STG) and the MGB were active when participants listened to sound (degraded and clear speech) compared with silent trials, confirming that the hemodynamic response to auditory stimulation was captured by the scans. L, left; R, right. Error bars indicate SEM.

176 sagittal slices, acquisition matrix of 256 × 240, voxel size = 1 mm^3].

In one participant, we were only able to acquire 136 (as opposed to 144) scans for Experiment 1 due to technical problems with cardiac gating. A second participant had to be excluded from all analyses concerning Experiment 2, because scan acquisition had become desynchronized with stimulus presentation.

Data analysis
Preprocessing. MRI data were analyzed in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Preprocessing was performed separately for Experiment 1 and 2. Structural MRI scans were manually aligned with the coordinate system such that AC–PC was in line with the *x*-axis and AC in the origin of the coordinate system. Functional images were realigned and unwarped using a field map, coregistered with the structural scan, segmented and normalized to standard space (Montreal Neurological Institute [MNI] space) using the segmentation parameters, and smoothed with an isotropic Gaussian kernel of 8 mm full-width at half-maximum (FWHM).

MR images were statistically analyzed in the context of the general linear model. We set up three different models for Experiment 1 and one model for Experiment 2 to assess the following effects:
Effects of auditory stimulation. In a basic model for Experiment 1, we defined three conditions at the single subject level: degraded speech, clear speech, and silent trials. The effect of auditory stimulation was tested by contrasting sound (degraded and clear speech) against silent trials. To avoid overspecification, silent trials were removed from all subsequent analyses.
Effects of speech degradation and trial-by-trial fluctuation with comprehension. In a speech-degradation model, we included two conditions: degraded and clear speech. Additionally, a parametric modulator of the degraded speech trials was defined, representing the behavioral report scores. A regressor of no interest, containing report latencies, was added to account for differences in speech production (analysis explained in detail below); this regressor was included in all remaining analyses. To assess effects of stimulus clarity, we contrasted degraded against clear speech trials. To reveal effects of trial-by-trial fluctuations in speech comprehension, we assessed correlations with the regressor representing report scores.
Effects of perceptual adaptation. To model effects of adaptation, we looked for signal changes over time corresponding to participants' slow performance increase ("adaptation curves"). However, there are a number of unspecific reasons why the BOLD signal could gradually change over time, for example, scanner drift or fatigue of the participant. Therefore, we compared the changes over time for vocoded speech to the change in activity seen for the clear speech condition, while taking into account the behaviorally observed adaptation to vocoded speech: rather than simply testing a condition × time interaction (which would not consider the adaptation curve of a listener), we tested for the three-way interaction for condition × time × behavior.

To this end, we created two parametric modulators, one for each condition, by multiplying time (i.e., trial number) with the linear adaptation curves. This resulted in a quadratic curve for the degraded speech regressor of which the slope was dependent on the individual adaptation curve (Fig. 5, top). However, since there was no perceptual adaptation in the clear speech condition (all sentences were fully comprehended), the "adaptation" curve was flat such that multiplication effectively left the time regressor unchanged, resulting in a linear curve.

Contrasts tested for the difference in these two regressors, ruling out the possibility that a slow change over time is due to slow unspecific signal drifts (which would be present in both conditions; analysis as in Büchel et al., 1998). Thus, we identified areas where changes over time are more pronounced in the degraded speech condition (where perceptual processing improves) than in the clear speech condition (where perceptual processing remains stable throughout).
Regressor of no interest for report latency. Although the present study was designed to image degraded speech perception, parts of the observed activity may be related to speech production or preparation, because participants overtly repeated back what they had understood starting ~3.5 s before scan acquisition (Fig. 1B). In particular, participants' verbal responses might have been faster for clear relative to degraded speech trials, perhaps leading to partly imaging the BOLD response to speech production, but more so for clear speech trials. Therefore, differences in report latencies might confound the comparison between degraded and clear speech trials. Similarly, adaptation-related signal changes might become confounded, as report latencies likely decrease as participants adapt to degraded speech. To control for this potential confound, we calculated report latency relative to the onset of the visual response cue (Fig. 1B). This measure was included at the first level as one single regressor of no interest in all models concerning Experiment 1 (described above). For trials where participants did not produce an overt response, the subject-specific mean report latency was entered.
Effects of ΔAM rate. In Experiment 2 we modeled four conditions; one for each level of AM rate difference between standard and deviant (referred to as "ΔAM rate"): ΔAM rate = 0 Hz, ±0.67 Hz, ±1.33 Hz, and ±2 Hz (note: in the ΔAM rate = 0 Hz condition, the deviant was modulated also at 4 Hz). To assess a linear correlation with ΔAM rate, these conditions were weighted with the contrast vector [−3 −1 1 3] for a positive correlation and [3 1 −1 −3] for a negative correlation with ΔAM rate. Large effects in these contrasts would mean linear scaling with deviance from the standard AM rate. In an additional conjunction analysis (Friston et al., 1999), we tested for the intersection of the effects of ΔAM rate and speech degradation.

All described analyses were whole brain analyses. Regressors were modeled using a finite impulse response comprising one bin. A high-pass filter with a cutoff of 1024 s was applied to eliminate low-frequency noise. No correction for serial autocorrelation was necessary because of the long TR in the sparse-sampling acquisition.

Second level statistics were calculated using a one-sample *t*-test. Group inferences are reported at a familywise error (FWE) corrected voxelwise threshold of $p < 0.05$, where FWE rate was controlled using random field theory. Only for the adaptation analyses did we use a slightly more lenient threshold of $p < 0.001$, where cluster-extent ($k > 20$) was corrected based on a Monte Carlo Simulation (Slotnick et al., 2003). *T*-statistic maps were transformed to *Z*-statistic maps using spm_t2z.m, and overlaid and displayed on the ch2 template in MNI space included with MRIcron (Rorden and Brett, 2000).
Lateralization analysis. Processes of comprehension are likely left-lateralized (Obleser et al., 2007; Rosen et al., 2011; McGettigan et al., 2012a). We therefore tested for lateralization of activity related to trial-by-trial fluctuation in speech comprehension. As in the analyses described above, EPI images first were realigned, unwarped, and coregistered. The images were then segmented using symmetric gray matter, white matter, and CSF templates, which were created by averaging each original template with itself flipped along the *y*-axis (Salmond et al., 2000). The segmentation parameters were used to normalize the images to MNI-space, resulting in a normalization to a symmetric MNI template (Liégeois et al., 2002) and smoothed at 8 mm FWHM. To conduct a voxel-by-voxel statistical test of laterality, the first level analyses were performed as described above. Resulting maps were flipped along the *y*-axis and compared with the original maps in a paired *t*-test (Bozic et al., 2010).
Regions of interest analyses. To extract measures of percentage signal change in the regions identified by the whole-brain analyses described above, we defined regions of interest (ROIs) using the SPM toolbox MarsBar (Brett et al., 2002). ROIs were defined as spheres with a radius of 3 mm centered on the identified peak coordinates. Voxels within an ROI were aggregated into a single contrast estimate using the first eigenvariate.

Results
Behavioral results
In Experiment 1, participants reported on average 51.9 ± 1.4% (mean ± SEM) words correctly per degraded sentence. Performance in clear trials was at 99.7 ± 0.2% correct.

We compared whether a linear or power law function (Fig. 2A) would better describe the report scores' increase over time by

calculating the BIC for each fit and each participant. The BIC scores for the linear fits (median 242.44, range 217.41–265.11) were smaller than those for power law fits (median 247.98, range 217.93–269.75), as shown by a Wilcoxon signed-rank test ($p <$ 0.001), indicating that the linear curve better fit the behavioral data. Thus, we chose the linear fit to describe the participants' improvement ("adaptation curve").

A one-tailed t-test on the slopes of the adaptation curves showed that they were significantly greater than zero ($t_{(29)} =$ 13.18, $p <$ 0.001), indicating that participants adapted to degraded speech, i.e., that their speech perception performance improved over the course of the experiment (Fig. 2A).

In Experiment 2, all participants performed well on the discriminable AM stimuli (mean $\pm$ SEM = 96.9 $\pm$ 0.7%). A paired samples t-test showed that there was no improvement from the first half (96.9 $\pm$ 0.7%) to the second half (96.7 $\pm$ 0.9%) of the experiment ($t_{(29)} =$ 0.34, $p =$ 0.74).

Controlling for speech production-related activations
To dissociate perceptual and response demands, we estimated report latencies. Report latency in degraded speech trials (977 $\pm$ 72 ms, mean $\pm$ SEM) was significantly longer than in clear speech trials (662 $\pm$ 30 ms; $t_{(29)} =$ 9.03, $p <$ 0.001). Moreover, report latency decreased over time, based on a t-test on the slopes of linear fits to report latencies as a function of trial ($-2.5 \pm 0.26 \times 10^{-3}$, mean $\pm$ SEM; $t_{(29)} =$ 9.94, $p <$ 0.001). Therefore, report latency was regressed out in all fMRI analyses. Importantly, analyses without this regressor of no interest yielded very similar fMRI activation patterns (results not reported). We take this as strong evidence that the observed effects are not driven by speech production, but rather perception.

fMRI results
We found extensive activation of the auditory system when contrasting sound against silent trials, in Experiment 1 (Heschl's gyrus, HG; planum temporale; superior temporal gyrus, STG; medial geniculate body, MGB; Fig. 2B) as well as in Experiment 2 (results not shown), confirming that the BOLD response to auditory stimulation was captured by scan acquisition.

Effects of physical speech degradation: degraded versus clear speech
To reveal regions that are modulated by physical speech degradation or clarity, we compared vocoded with clear speech trials. Areas where degraded relative to clear speech yielded an increased BOLD signal included the left supplementary motor area/ anterior cingulate cortex (SMA/ACC), anterior insula, and caudate nucleus bilaterally. In contrast, clear compared with degraded speech yielded higher activations bilaterally in the precentral gyrus spanning the temporal cortices, supramarginal gyrus (SMG), putamen, posterior cingulate cortex, and angular gyrus bilaterally (Fig. 3A, Table 1).

AM rate discrimination and degraded speech processing
The magnitude of the AM rate difference between standard and deviant stimuli (ΔAM rate) correlated positively with activity in right HG, left amygdala, and SMG (Fig. 3B; Table 1), signifying that activity increased with larger deviance from the standard (and thus easier AM rate discrimination). Conversely, ΔAM rate correlated negatively with activity in the ACC/SMA, left insula, and IFG, meaning that the signal increased as modulation rate differences diminished and discrimination became more difficult.

In a series of conjunction analyses, we identified regions commonly involved in processing of degraded speech and of AM stimuli: a conjunction analysis between (1) positive correlation with ΔAM rate and (2) clear > degraded speech was significant in the SMG bilaterally and posterior cingulate. A conjunction of the inverse contrasts (negative correlation with ΔAM rate $\cap$ degraded > clear speech) yielded a significant cluster in the SMA/ ACC, insula bilaterally, and left IFG (Table 1). There were no commonly activated areas for the "cross-over" conjunction "negative correlation with ΔAM rate $\cap$ clear > degraded speech" or vice versa.

Trial-by-trial fluctuation with speech comprehension
To reveal areas reflecting trial-by-trial fluctuations in speech comprehension, we tested for correlations with the behavioral report scores. The BOLD signal linearly increased with improved comprehension of degraded speech in bilateral temporal cortices comprising HG, superior temporal sulcus (STS), left IFG, precentral gyrus bilaterally, the putamen, the thalamus including MGB bilaterally, left angular gyrus, frontal medial cortex, posterior cingulate, and cerebellum (Fig. 4). There were no negative correlations between the fMRI signal and report scores. An additional voxel-by-voxel laterality analysis of this speech comprehension effect (see Materials and Methods) revealed that activity in the angular gyrus was significantly left-lateralized (Table 1). This is apparent in Figure 4, where angular gyrus activity is seen in the sagittal slice of the left hemisphere only.

Effects of perceptual adaptation
To identify brain areas that change over time as a function of adaptation, we compared changes over time between degraded and clear speech while taking into account the behaviorally observed adaptation, i.e., we tested the time $\times$ condition $\times$ behavior interaction (Fig. 5, top). Thus, we separated changes over time due to adaptation (which only occurred in the degraded speech) from unspecific slow signal changes (which would be present in both conditions and hence be canceled out in the contrast). As the behavioral change is marked for vocoded speech but virtually absent for clear speech, the BOLD signal is expected to change at different rates (i.e., more change in the vocoded condition because, on top of scanner drift, adaptation-related signal changes are present). Therefore, we modeled a more pronounced increase in the vocoded (quadratic increase) than in the clear speech condition (linear increase). This model separates uninteresting signal drifts from the effects due to the behaviorally observed adaptation.

At a threshold of p(FWE) < 0.05, we found stronger downregulation over time in degraded relative to clear speech trials as listeners adapted in the anteroventral thalamic nucleus (Morel, 2007). At the slightly more lenient threshold of $p <$ 0.001 (cluster-extent corrected), the caudate, frontal regions and an occipital cluster spanning from the cerebellum to fusiform gyrus up to the precuneus showed the same effect. Conversely, an activity increase over time in degraded more than clear speech with adaptation was found in the left precentral gyrus and posterior cingulate cortex (Fig. 5; Table 1).

Discussion
The present study was designed to reveal neural systems that support rapid perceptual adaptation to degraded speech and contributes three major novel findings. First, degraded more than clear speech activates an "executive" system, which overlaps with the neural substrates of difficult auditory discrimination, namely
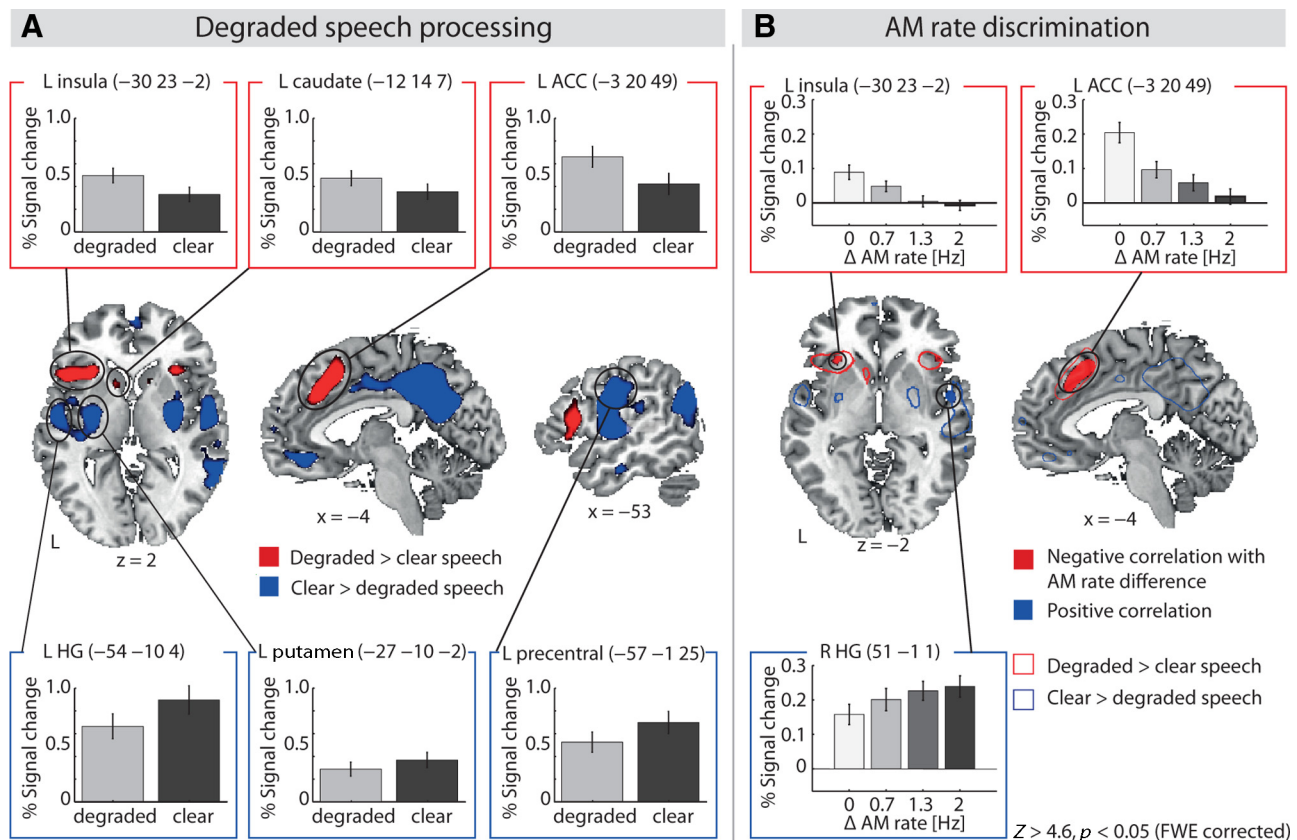
Figure 3. *A*, Areas sensitive to speech degradation. Degraded relative to clear speech yielded increased activation in the anterior insula, the caudate, the SMA/ACC, and IFG (in red). Conversely, clear compared with degraded speech revealed increased activity in the temporal cortices, putamen, posterior cingulate, precentral gyrus, SMG and angular gyrus (AG). Estimated percentage signal change for vocoded and clear trials is shown for selected regions. *B*, Effects of ΔAM rate in Experiment 2 and overlaps with Experiment 1. As AM rate difference (ΔAM rate) between standard and deviant increased, activity reduced in the SMA/ACC and insula (red). Activity increased with higher ΔAM rate in right Heschl's gyrus, left SMG and amygdala (blue). The contours of the Z-statistic maps showing effects of speech degradation are overlaid to illustrate overlaps in activation. A conjunction analysis between the effects of negative correlation with ΔAM rate and degraded > clear speech was significant in the SMA/ACC, insula bilaterally, and the IFG; conjunction of the inverse contrasts (positive correlation with ΔAM rate ∩ clear > degraded speech) reached significance in the posterior cingulate and SMG bilaterally. L, left; R, right. Error bars indicate SEM.

in the insula and SMA/ACC. Second, BOLD signal in a "language" network, comprising auditory, premotor cortices, and left angular gyrus, depended on speech comprehension rather than physical clarity of the stimulus. Third, the data provide first evidence that self-regulated perceptual adaptation to degraded speech co-occurs with a BOLD downregulation of subcortical structures.

Executive mechanisms in speech and non-speech processing
Degraded more than clear speech evoked enhanced activity in anterior areas including the insula and the SMA/ACC. This has been observed consistently for difficult comprehension (Giraud et al., 2004; Eckert et al., 2009). Clear speech, in contrast, revealed an expected activity increase in the bilateral temporal cortices (Scott et al., 2000; Davis and Johnsrude, 2003; Giraud et al., 2004; Wild et al., 2012a).

When examining to which extent difficult speech and non-speech perception rely on joint neural substrates, a conjunction revealed SMA/ACC and the anterior insula bilaterally. This comparison pertains directly to contributions of bottom-up (i.e., fidelity of AM representations in the ascending auditory pathway) versus top-down mechanisms (e.g., attentional processes) in degraded speech perception. For example, frequency discrimination learning shows partial specificity to the trained frequency (Amitay et al., 2006), consistent with a bottom-up account. In contrast, top-down mechanisms of attention are plausibly in-

volved in perceptual learning (Halliday et al., 2011). Training improves the ability to attend to a task-specific stimulus dimension, and discrimination learning occurs even in the absence of a discriminable stimulus difference (Amitay et al., 2006). Similarly, speech degradation studies have found evidence for bottom-up accounts (Sebastián-Gallés et al., 2000; Hervais-Adelman et al., 2008; Idemaru and Holt, 2011) as well as top-down accounts of perceptual learning, in which lexical information aids perceptual adaptation (Davis and Johnsrude, 2003; Davis et al., 2005).

The structures commonly recruited for degraded speech processing and AM discrimination are clearly not specific to auditory envelope processing, but have been suggested by a number of studies to be involved rather in top-down, executive processes (Adank, 2012). Eckert et al. (2009) demonstrated that the insula and SMA/ACC are engaged when tasks become increasingly difficult, independent of modality or task, suggesting that these regions subserve executive processes. Consistently, the anterior insula and SMA showed an enhanced BOLD signal when listeners attended to speech (rather than a distracter), and the speech signal was increasingly degraded (vocoded rather than clear; Wild et al., 2012b). More specifically, these regions are a resource for attention and performance monitoring processes (Dosenbach et al., 2006, 2007; Sadaghiani and D'Esposito, 2012). Thus, we argue that the anterior insula and SMA/ACC fulfill executive processes, and that the recruitment of these executive components is pivotal for a wide range of challenging listening situations.

Table 1. Overview of MRI activation clusters, thresholded at voxel-wise p(FWE) < 0.05, unless indicated differently

Location	MNI coordinates			Z-Score	Extent (mm³)
Degraded > clear speech					
L SMA/ACC	−6	26	40	7.19	3060
L anterior insula	−30	20	−5	7.05	3825
R anterior insula	33	23	−3	5.84	576
L caudate nucleus	−9	11	7	5.18	162
R caudate nucleus	12	17	10	5.45	171
Clear > degraded speech					
R superior frontal gyrus	6	59	28	6.55	7866
R precentral gyrus	54	−7	16	6.5	28647
R putamen	27	−10	−2	7.19	
R MTG	63	−19	−11	6.79	
R angular gyrus	51	−58	22	7.94	
L precentral gyrus	−54	−1	22	7.59	24183
L putamen/globus pallidus	−27	−10	−2	6.87	
L HG	−54	−10	7	7.18	
L angular gyrus	−51	−67	34	7.24	
L MTG	−60	−13	−20	6.69	1044
L posterior cingulate	−6	−37	43	7.47	20700
Positive correlation with ΔAM rate					
R HG	51	−1	1	5.95	270
L amygdala	−24	−7	−14	5.37	135
L SMG	−57	−31	19	5.17	279
Negative correlation with ΔAM rate					
L SMA/ACC	−6	17	49	6.19	1143
L anterior insula	−30	23	−2	4.85	54
L IFG, pars opercularis	−48	14	28	5.16	270
Positive correlation with ΔAM rate ∩ clear > degraded speech					
L SMG	−57	−28	22	5.22	477
R SMG	60	−25	19	5.08	261
L posterior cingulate cortex	0	−40	46	5.14	486
Negative correlation with ΔAM rate ∩ degraded > clear speech					
L SMA/ACC	−6	17	49	7.15	1602
L anterior insula	−30	23	−2	5.65	324
R anterior insula	33	23	1	5.25	360
L IFG, pars opercularis	−48	14	25	5.17	315
Positive correlation with trial-by-trial fluctuations in comprehension					
L frontal medial cortex	−3	50	−14	6.62	765
L IFG	−51	26	13	5.75	243
L STS	−57	−7	−20	7.47	18279
L precentral gyrus	−54	−10	34	6.58	
L HG	−57	−13	4	6.67	
L putamen	−30	−16	−2	6.21	
L angular gyrus	−51	−61	19	7.50	
R precentral sulcus	57	−4	31	6.74	11475
R HG	57	−13	4	6.7	
R putamen	30	−7	−8	6.34	
L thalamus	−12	−19	4	5.56	885
R thalamus	12	−19	4	5.57	306
L posterior cingulate	−3	−52	16	6.22	2376
R cerebellum	15	−61	−23	6.08	3141
L cerebellum	−12	−64	−23	5.73	
Lateralization of comprehension network					
L angular gyrus	−45	−70	37	5.83	378
Decrease with time × behavior for degraded > clear speech, $p < 0.001$, $k > 20$					
L superior frontal gyrus	−27	47	31	3.79	666
R MFG	33	47	22	4.89	729
R anteroventral thalamus	3	−4	7	5.07	7929
R caudate	18	20	−5	4.48	
L fusiform gyrus	−24	−61	−14	4.84	19116
Increase with time × behavior for degraded > clear speech, $p < 0.001$, $k > 20$					
L precentral gyrus	−54	−1	19	3.8	189
Posterior cingulate	0	−58	25	4.05	333

L, left; R, right; ACC, anterior cingulate cortex; HG, Heschl's gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; SMA, supplementary motor area; SMG, supramarginal gyrus; STG, superior temporal gyrus; STS: superior temporal sulcus.

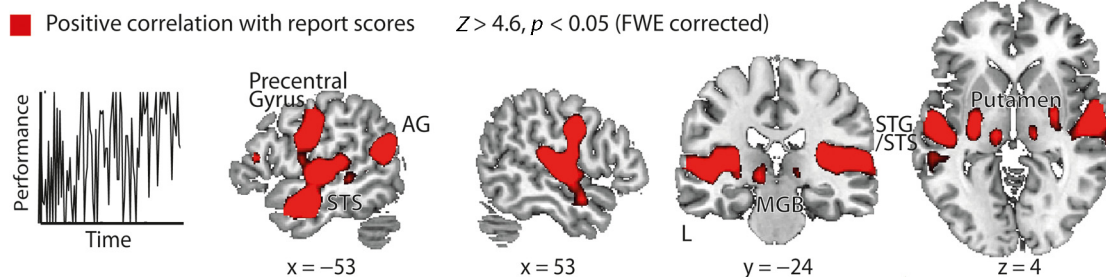## Trial-by-trial fluctuations in degraded speech comprehension

Figure 4. Areas varying with speech comprehension. With better speech comprehension on a given trial (graph shown on the left), activity increased in a network comprising the auditory cortices, left IFG, precentral gyrus, MGB, putamen, angular gyrus (AG), and cerebellum. There were no negative correlations with trial-by-trial report scores.
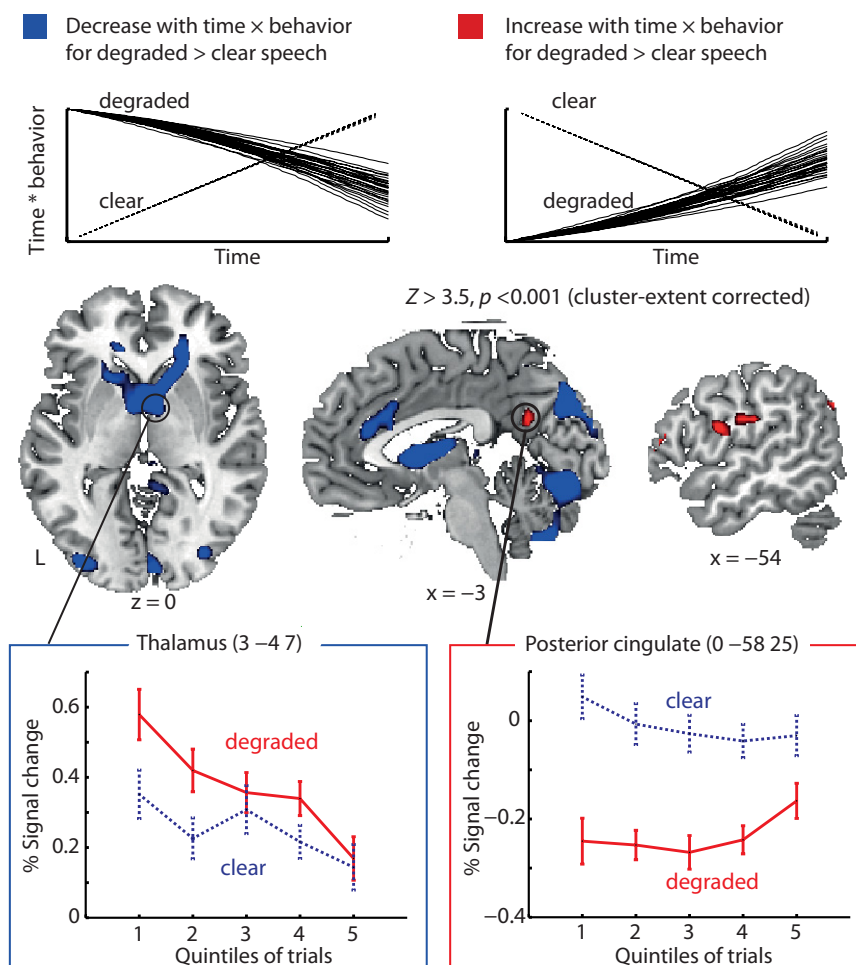

Figure 5. Effects of adaptation. To detect changes in activity due to adaptation, we compared the changes over time with performance increase between degraded and clear speech (top). Stronger downregulation for degraded relative to clear speech was observed in the anteroventral thalamic nucleus, caudate, frontal, occipital, and cerebellar regions (blue). Conversely, activity in the posterior cingulate and left ventral premotor cortex increased over time with adaptation (middle). To illustrate changes over time in the thalamus and posterior cingulate, we split up the experiment *post hoc* into quintiles for which we separately calculated the contrast estimates for each condition (bottom).

Revisiting the speech comprehension network

A number of earlier imaging studies manipulated physical stimulus features to vary speech intelligibility and found sensitivity to these manipulations along the superior temporal gyrus and sul-

cus, often extending into prefrontal and inferior parietal regions (Scott et al., 2000; Davis and Johnsrude, 2003; Zekveld et al., 2006; Obleser et al., 2007; Obleser and Kotz, 2010). In contrast, the present study held physical stimulus properties constant (i.e., 4-band vocoding). Therefore, we were able to identify regions where activation varied with actual speech comprehension (i.e., behavioral report scores), independent of acoustic differences.

For a given trial, activity in a perisylvian network was tightly coupled to comprehension. Importantly, these areas (Fig. 4) overlapped largely with those activated by clear compared with degraded speech (Fig. 3A, blue Z-statistic maps). Together, this is strong evidence that the observed network supports sensorimotor operations involved in successful comprehension (and, hence, successful repetition) rather than simply indexing sensitivity to physical stimulus characteristics.

Such linguistic processes of comprehension have been proposed to be largely processed in the left hemisphere (McGettigan et al., 2012a; Peelle, 2012). Concordantly, we found evidence for a left-lateralization of activity in the angular gyrus, a structure that is associated with semantic processing (Ferstl and von Cramon, 2002; for review see Price, 2012) and has been suggested to facilitate speech comprehension when signal quality declines (Obleser et al., 2007; Obleser and Kotz, 2010).

Cortical contributions to perceptual adaptation

One important objective here was to identify and track on-line the neural systems supporting adaptation to degraded speech. Depending on adaptation, activity increased over time in cortical areas of the premotor cortex and posterior cingulate and decreased in frontal and occipital areas.

The premotor cortex has been suggested to mediate successful perceptual learning of degraded speech by mapping the unfamil-

iar auditory signal onto existing articulatory representations of speech sounds. Consistent with this idea, adaptation to time-compressed speech is reflected in a downregulation of activity in the ventral premotor cortex (Adank and Devlin, 2010). Hervais-Adelman et al. (2012) observed precentral sulcus activation, when a vocoded word was paired with its clear representation, a type of feedback that is known to enhance perceptual learning (Davis et al., 2005). Similarly, we note in the present study, although with feedback-free learning, that activity in the premotor cortex increases as a listener adapts to vocoded speech. Thus, perceptual learning might be rooted in sensorimotor integration mediated by speech-productive areas.

Enhanced activity in the posterior cingulate cortex has been noticed by Obleser et al. (2007) when semantic context helped comprehension of degraded speech at an intermediate signal quality (8-band vocoded speech). This is commensurate with the current data where posterior cingulate activity not only correlated with successful speech comprehension at a given trial (Table 1), but increased as a listener adapted to degraded speech. Hence, adaptation-related increase of posterior cingulate activity might directly relate to its facilitating role in degraded speech comprehension.

Although the present experiment did not involve visual manipulations, we found occipital cortex activity (fusiform gyrus, precuneus) to correlate negatively with adaptation. While the fusiform gyrus has been implicated in audiovisual speech perception (Stevenson et al., 2010; Nath and Beauchamp, 2011; McGettigan et al., 2012b; von Kriegstein et al. (2003) observed modulation of fusiform gyrus activity during auditory speech perception, even in the absence of visual stimuli. Similarly, Giraud et al. (2001) have established that visual cortex contributes to auditory speech processing in CI patients but also in normal-hearing listeners, especially when listening to meaningful sentences (Giraud and Truy, 2002). They hypothesized that auditory-to-visual cross-modal interaction contributes to semantic processing. In line with these studies, we speculate that initial recruitment of the visual cortex might help a listener extract meaning when first confronted with a novel form of speech degradation.

A previous study, which used a feedback-based vocoded-speech learning paradigm simulating CI rehabilitation programs (Eisner et al., 2010), found the IFG to be involved in successful adaptation. The authors attributed to the IFG a role in the "specific use of simultaneous written feedback to enhance comprehension," in line the view of the IFG serving as integration site for different sources of information necessary for speech comprehension (Hagoort, 2005; Rauschecker and Scott, 2009). In the present study, we did not observe IFG involvement in adaptation. Likely due to the absence of feedback, listeners might have relied on substantially different neural mechanisms to adapt to degraded speech.

Subcortical contributions to perceptual adaptation

A major novel contribution of the present study was the detection of subcortical involvement in perceptual adaptation. In previous imaging studies, such subcortical contributions might have been obscured due to heartbeat-related artifacts, which we avoided by use of a cardiac-gated scanning protocol.

We identified the anteroventral nucleus of the thalamus to be functionally involved in adaptation (Fig. 5); note that this structure is proximal to but not identical to the pulvinar, where structural differences have been reported as predictive of degraded speech adaptation before (Erb et al., 2012). Adaptation-related decrease in activity also encompassed the caudate. Although the basal ganglia have primarily been implicated in motor function, there is accumulating evidence that they play an important role in language processing (Lieberman et al., 1992; Kotz et

al., 2002; Fiebach et al., 2003; Kotz, 2006). Presumably, they exert their function in language processing through their high connectivity to the cortex (Crosson, 1999): The anteroventral thalamic nucleus and the caudate form part of the cortico-striato-thalamo-cortical loop, which is proposed to collect cortical information and funnel and converge it at cortical output areas, thereby reconfiguring cortical activation patterns (Kemp and Powell, 1970; O'Connell et al., 2011). In the context of adaptation to degraded speech, naive listeners might at first be forced to rely more on information-selection processes supported by the cortico-striato-thalamo-cortical loop. Engagement of this pathway is likely to sharpen the cortical representation of a stimulus and ultimately lead to a convergence of the degraded speech signal onto a clear-speech "representation," allowing for enhanced comprehension.

Conclusions

The present work elucidates the central neural mechanisms of rapid adaptation to acoustic speech degradation with respect to three points. First, when listening tasks become increasingly difficult, in the speech as well as the non-speech domain, listeners rely on a common executive network for "effortful listening" (Eckert et al., 2009), involving the SMA/ACC and anterior insula. Second, a perisylvian network subserves speech comprehension and fluctuates with actual comprehension rather than physical stimulus features. Finally, the present data advance the understanding of how a listener adapts to a degraded speech input, demonstrating that rapid adaptation is partly explained by hemodynamic downregulation in subcortical structures.

References

Adank P (2012) The neural bases of difficult speech comprehension and speech production: two Activation Likelihood Estimation (ALE) meta-analyses. Brain Lang 122:42–54. CrossRef Medline

Adank P, Devlin JT (2010) On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. Neuroimage 49:1124–1132. CrossRef Medline

Amitay S, Irwin A, Moore DR (2006) Discrimination learning induced by training with identical stimuli. Nat Neurosci 9:1446–1448. CrossRef Medline

Bent T, Buchwald A, Pisoni DB (2009) Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. J Acoust Soc Am 126:2660–2669. CrossRef Medline

Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD (2010) Bihemispheric foundations for human speech comprehension. Proc Natl Acad Sci U S A 107:17439–17444. CrossRef Medline

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox. Presentation at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, Sendai, Japan.

Büchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron 20:947–957. CrossRef Medline

Crosson B (1999) Subcortical mechanisms in language: lexical-semantic mechanisms and the thalamus. Brain Cogn 40:414–438. CrossRef Medline

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431. Medline

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 134:222–241. CrossRef Medline

Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. Neuron 50:799–812. CrossRef Medline

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A 104:11073–11078. CrossRef Medline

Eckert MA, Menon V, Walczak A, Ahlstrom J, Denslow S, Horwitz A, Dubno

JR (2009) At the heart of the ventral attention system: the right anterior insula. Hum Brain Mapp 30:2530–2541. CrossRef Medline

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 30:7179–7186. CrossRef Medline

Erb J, Henry MJ, Eisner F, Obleser J (2012) Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. Neuropsychologia 50:2154–2164. CrossRef Medline

Fallon JB, Irvine DR, Shepherd RK (2008) Cochlear implants and brain plasticity. Hear Res 238:110–117. CrossRef Medline

Ferstl EC, von Cramon DY (2002) What does the frontomedian cortex contribute to language processing: coherence or theory of mind? Neuroimage 17:1599–1612. CrossRef Medline

Fiebach CJ, Friederici AD, Müller K, von Cramon DY, Hernandez AE (2003) Distinct brain representations for early and late learned words. Neuroimage 19:1627–1637. CrossRef Medline

Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. Neuroimage 10:385–396. CrossRef Medline

Giraud AL, Truy E (2002) The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. Neuropsychologia 40:1562–1569. CrossRef Medline

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. J Neurophysiol 84:1588–1598. Medline

Giraud AL, Price CJ, Graham JM, Truy E, Frackowiak RS (2001) Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron 30:657–663. CrossRef Medline

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. CrossRef Medline

Golestani N, Zatorre RJ (2004) Learning new sounds of speech: reallocation of neural substrates. Neuroimage 21:494–506. CrossRef Medline

Greenwood DD (1990) A cochlear frequency-position function for several species–29 years later. J Acoust Soc Am 87:2592–2605. CrossRef Medline

Hagoort P (2005) On Broca, brain, and binding: a new framework. Trends Cogn Sci 9:416–423. CrossRef Medline

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223. CrossRef Medline

Halliday LF, Moore DR, Taylor JL, Amitay S (2011) Dimension-specific attention directs learning and listening on auditory training tasks. Atten Percept Psychophys 73:1329–1335. CrossRef Medline

Harris KC, Dubno JR, Keren NI, Ahlstrom JB, Eckert MA (2009) Speech recognition in younger and older adults: a dependency on low-level auditory cortex. J Neurosci 29:6078–6087. CrossRef Medline

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 34:460–474. CrossRef Medline

Hervais-Adelman A, Carlyon RP, Johnsrude I, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. Lang Cogn Proc 27:1145–1166. CrossRef

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline

Idemaru K, Holt LL (2011) Word recognition reflects dimension-based statistical learning. J Exp Psychol Hum Percept Perform 37:1939–1956. CrossRef Medline

Kalikow DN, Stevens KN, Elliott LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 61:1337–1351. CrossRef Medline

Kemp JM, Powell TP (1970) The cortico-striate projection in the monkey. Brain 93:525–546. CrossRef Medline

Kotz SA (2006) The role of the basal ganglia in auditory language processing: evidence from ERP lesion studies and functional neuroimaging. Leipzig: MPI Series in Human Cognitive and Brain Sciences.

Kotz SA, Cappa SF, von Cramon DY, Friederici AD (2002) Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. Neuroimage 17:1761–1772. CrossRef Medline

Lieberman P, Kako E, Friedman J, Tajchman G, Feldman LS, Jiminez EB

(1992) Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. Brain Lang 43:169–189. CrossRef Medline

Liégeois F, Connelly A, Salmond CH, Gadian DG, Vargha-Khadem F, Baldeweg T (2002) A direct test for lateralization of language activation using fMRI: comparison with invasive assessments in children with epilepsy. Neuroimage 17:1861–1867. CrossRef Medline

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK (2012a) An application of univariate and multivariate approaches in FMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. J Cogn Neurosci 24:636–652. CrossRef Medline

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK (2012b) Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia 50:762–776. CrossRef Medline

Morel A (2007) Stereotactic atlas of the human thalamus and basal ganglia. New York: Informa Healthcare.

Nath AR, Beauchamp MS (2011) Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J Neurosci 31:1704–1714. CrossRef Medline

Obleser J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. Cereb Cortex 20:633–640. CrossRef Medline

Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289. CrossRef Medline

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805–817. CrossRef Medline

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113. CrossRef Medline

Peelle JE (2012) The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Front Hum Neurosci 6:309. Medline

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 31:1315–1330. CrossRef Medline

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387. CrossRef Medline

Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62:816–847. CrossRef Medline

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. CrossRef Medline

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behav Neurol 12:191–200. Medline

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 106:3629–3636. CrossRef Medline

Rosen S, Wise RJ, Chadha S, Conway EJ, Scott SK (2011) Hemispheric asymmetries in speech perception: sense, nonsense and modulations. PloS One 6:e24672. CrossRef Medline

Sadaghiani S, D'Esposito M (2012) Dissociating brain networks for alertness and selective attention. Soc Neurosci Abstr 38.492.13.

Salmond CH, Ashburner J, Vargha-Khadem F, Gadian DG, Friston KJ (2000) Detecting bilateral abnormalities with voxel-based morphometry. Hum Brain Mapp 11:223–232. CrossRef Medline

Schwarz G (1978) Estimating the dimension of a model. Ann Statist 6:461–464. CrossRef

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406. CrossRef Medline

Sebastián-Gallés N, Dupoux E, Costa A, Mehler J (2000) Adaptation to time-compressed speech: phonological determinants. Percept Psychophys 62:834–842. CrossRef Medline

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304. CrossRef Medline

Slotnick SD, Moo LR, Segal JB, Hart J Jr (2003) Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Brain Res Cogn Brain Res 17:75–82. CrossRef Medline

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of

asynchronous audiovisual speech perception. Neuroimage 49:3308–3318. CrossRef Medline

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48–55. CrossRef Medline

von Kriegstein K, Patterson RD, Griffiths TD (2008) Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. Curr Biol 18:1855–1859. CrossRef Medline

Wild CJ, Davis MH, Johnsrude IS (2012a) Human auditory cortex is sensi-

tive to the perceived clarity of speech. Neuroimage 60:1490–1502. CrossRef Medline

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012b) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021. CrossRef Medline

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