

Changing meaning causes coupling changes within higher levels of the cortical hierarchy

T. M. Schofield^{a,1}, P. Iverson^b, S. J. Kiebel^c, K. E. Stephan^d, J. M. Kilner^a, K. J. Friston^a, J. T. Crinion^a, C. J. Price^a, and A. P. Leff^a

^aWellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom; ^bDivision of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom; ^cMax Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; and ^dLaboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, CH-8006 Zurich, Switzerland

Edited by Ranulfo Romo, Universidad Nacional Autónoma de México, Mexico, D.F., Mexico, and approved May 26, 2009 (received for review November 12, 2008)

Processing of speech and nonspeech sounds occurs bilaterally within primary auditory cortex and surrounding regions of the superior temporal gyrus; however, the manner in which these regions interact during speech and nonspeech processing is not well understood. Here, we investigate the underlying neuronal architecture of the auditory system with magnetoencephalography and a mismatch paradigm. We used a spoken word as a repeating “standard” and periodically introduced 3 “oddball” stimuli that differed in the frequency spectrum of the word’s vowel. The closest deviant was perceived as the same vowel as the standard, whereas the other 2 deviants were perceived as belonging to different vowel categories. The neuronal responses to these vowel stimuli were compared with responses elicited by perceptually matched tone stimuli under the same paradigm. For both speech and tones, deviant stimuli induced coupling changes within the same bilateral temporal lobe system. However, vowel oddball effects increased coupling within the left posterior superior temporal gyrus, whereas perceptually equivalent nonspeech oddball effects increased coupling within the right primary auditory cortex. Thus, we show a dissociation in neuronal interactions, occurring at both different hierarchical levels of the auditory system (superior temporal versus primary auditory cortex) and in different hemispheres (left versus right). This hierarchical specificity depends on whether auditory stimuli are embedded in a perceptual context (i.e., a word). Furthermore, our lateralization results suggest left hemisphere specificity for the processing of phonological stimuli, regardless of their elemental (i.e., spectrotemporal) characteristics.

dynamic causal modeling | language | magnetoencephalography | mismatch negativity | predictive coding

For speech to be understood, it is necessary for invariant, meaningful representations to be abstracted from an inconstant and time-varying acoustical signal. The basic symbolic units of speech are referred to as phonemes. These can be thought of as the units that differentiate one word from another (1). Different languages may use different sets of phonemes; 2 acoustically different speech sounds might convey a change of meaning in one language, but might not in another. Neuronal representations of these units necessarily accumulate information over time. They need to be accurate enough to identify the differences between 2 similarly sounding words (e.g., pin/pen) but flexible enough to accommodate variable, context-sensitive acoustic features imparted to the speech signal by different speakers (e.g., gender and age differences). In this study, we used a mismatch paradigm, magnetoencephalography (MEG) and dynamic causal modeling (DCM), to investigate how these symbolic representations are reflected in neuronal processing in auditory cortical areas.

The mismatch negativity (MMN) and its magnetic equivalent the mismatch field (MMF), together the mismatch response, are components of the event-related potential or field elicited by the auditory presentation of rare “oddball” or “deviant” stimuli within

a train of frequently presented “standard” stimuli. The mismatch response typically peaks between 150 and 250 ms after stimulus presentation and may be elicited in the absence of stimulus-directed attention (2–4). Within the auditory cortex, the response is thought to be generated within a network of 4 sources; these are the primary auditory cortices and posterior superior temporal gyri (STG) in the left and right hemispheres (5–8).

It has been proposed that the mismatch response can be explained in terms of a predictive coding framework (9). According to this view, with each repetition of the standard stimulus, the auditory system adjusts a generative model of the stimulus train to allow an increasingly precise prediction of the subsequent stimulus. Under this account, the mismatch response that occurs when the novel deviant stimulus is presented represents a failure of higher-level, top-down processes to suppress prediction error in lower-level auditory regions. This makes the mismatch response a useful tool for probing representations in the auditory system, because the characteristic of the auditory stimulus that causes the prediction error must be represented in the brain at this level of auditory processing. The results of previous mismatch experiments suggest that the abstract stimulus property of “native phoneme” is somehow represented in the brain during early cortical processing, because speech sound deviants belonging to a different native phoneme category from the standard elicit a greater mismatch response than speech deviants that do not (4, 10, 11).

However, the nature of this interaction between the “bottom-up” physical acoustic (i.e., the spectrotemporal profile) and the “top-down” linguistic properties (i.e., whether the sound belongs to a native phoneme category or not) of the stimuli is not well understood. In this article, we use MEG and DCM (12) to make inferences about the interaction of bottom-up and top-down processes during the processing of meaningful speech sounds. This technique enabled us to pinpoint where in the auditory processing hierarchy (and in which hemisphere) specific attributes (spectrotemporal and phonemic) of acoustic stimuli are processed.

To compare speech and nonspeech processing, we used 2 types of auditory stimuli matched for their psychoacoustic properties. The first type, vowel sounds, have long-standing neuronal representations built up over years of sensory experience. We chose

Author contributions: T.M.S., J.T.C., C.J.P., and A.P.L. designed research; T.M.S. and A.P.L. performed research; P.I., S.J.K., K.E.S., and J.M.K. contributed new reagents/analytic tools; T.M.S. and S.J.K. analyzed data; and T.M.S., P.I., S.J.K., K.E.S., J.M.K., K.J.F., J.T.C., C.J.P., and A.P.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: t.schofield@fil.ion.ucl.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811402106/DCSupplemental.

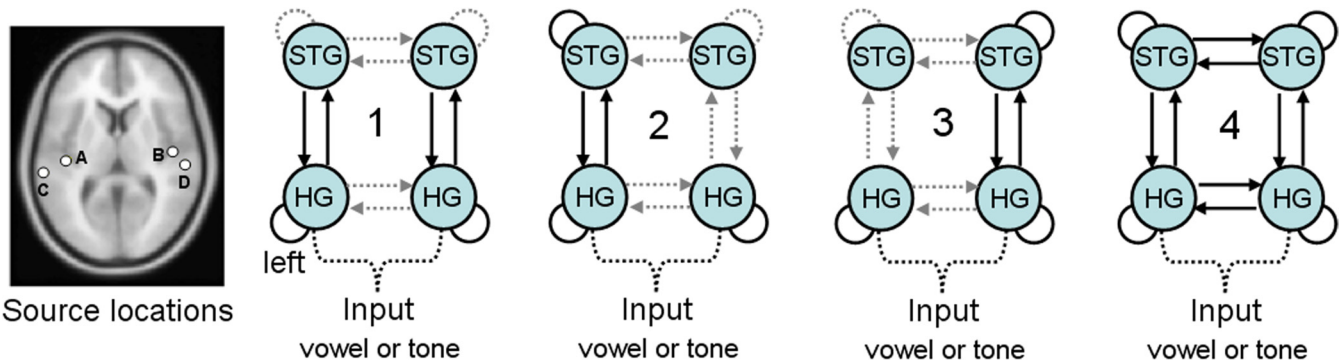


Fig. 1. All models contained 4 neuronal sources (far left). A, the left HG [MNI coordinates (−42, −22, 7)]; B, the right HG (46, −14, 8); C, the left STG (−61, −32, 8); D, the right STG (59, −25, 8). The 4 models used in the DCM analysis were, from left-to-right: 1) the initial model derived from Garrido et al. (15); 2) left-sided model; 3) right-sided model; 4) fully connected model. Model 4 explained the observed data (comprising 3 deviant ERFs) better than the other models, across both stimulus classes. Connections between regions that are not allowed to be modulated by differences in auditory stimuli (standards and deviants) are dashed (gray); this includes inputs into both primary auditory cortices (dashed black). Connections between regions that are allowed to vary between standard and deviant stimuli are shown as solid black lines. This characteristic differentiates the 4 basic models.

this type of phoneme because vowel sounds are distinguished by differences in their first and second formants (i.e., frequency peaks caused by resonances in the vocal tract) (13). We varied the formant frequencies of the vowel component of the standard stimulus to produce 3 deviant vowel sounds that were perceptually different from the standard. The vowels were embedded in short “consonant–vowel–consonant” words that differed only in the vowel sound. The standard word was “Bart.” The closest deviant (D1) differed sufficiently from the standard in the frequency of the first and second formants to be distinguishable, but was perceived as containing the same phonemes as the standard; that is, it sounded like the same word (Bart), spoken in a different manner. The other 2 deviants (D2 and D3) were further away acoustically from the standard and were perceived as different phonemes, thus forming different words (“Burt” and “beat,” respectively). We then generated a second class of stimuli, sinusoid tones (tone stimuli) centered on the second formant frequencies of the vowel stimuli (see *Materials and Methods* and *Fig. S1* for details), that do not have high-level neuronal representations. In line with previous studies, we expected (i) all speech and nonspeech deviants to produce mismatch responses, (ii) the amplitude of the mismatch would increase with deviancy (i.e., amplitude for the most distant deviant, D3, would be greater than for the least distant deviant, D1), and (iii) an interaction between deviancy and the nature of the stimuli (speech vs. nonspeech). In other words, the mismatch response (prediction error) would be greater for vowel D2 than tone D2 because the former enables more precise predictions based on phoneme category. By examining the mismatch responses elicited by these stimuli we could model the functional anatomy of preattentive processing of acoustic and phonemic speech sounds, within auditory cortex and associate any deviant-dependent differences with the spectrotemporal aspects of the stimuli (oddball vs. standard) and their perceptual context (word vs. tone).

Although the effects of speech sound deviancy on neuronal connectivity have not been reported, the simpler case of tone-frequency deviants and their effects on interaction within the auditory system have already been investigated in some detail by using DCM (14–16). Within auditory cortex, the mismatch response, measured with EEG, is best explained by an increase in postsynaptic sensitivity within each primary auditory region and increases in the influence of reciprocal connection (i.e., in both forward and backward connections) between the primary auditory region and ipsilateral STG (14). This architecture fits well with a predictive coding account of the mismatch response;

auditory input is received by the primary auditory regions and compared with the predictions generated by cortical processing in the higher levels of the auditory network. These predictions are finessed by repeated exposure to the standard stimulus such that when the incoming stimulus violates these predictions, it evokes a greater prediction error and a more exuberant exchange of signals between high and low areas (14, 17). The present study applied DCM and used Bayesian model selection (18) to determine the most likely location of changes in neuronal interactions that underlie oddball effects in phonemic processing. We were specifically hoping to show that changes in vowels involve changes in distributed processing at higher levels of the auditory hierarchy, relative to changes in tones. Furthermore, we predicted that vowel-dependent changes would show a left lateralization.

Neuronal responses were investigated by using DCM in 2 steps: First, we optimized the architecture of the network that could explain event-related fields (ERFs) for deviant stimuli. This process entailed the creation of a series of within-subject DCMs in which the effects of each deviant was modeled with changes in neuronal coupling in model-specific sets of connections. The relative evidence for each model was pooled over subjects to identify the best model. Second, at the between-subject level, we assessed coupling changes using the corresponding parameter estimates of the best model and simple *t* tests.

We tested 4 models (see Fig. 1 for a graphical summary). All models were based on 4 neuronal sources; the left primary auditory cortex (left HG), the right primary auditory cortex (right HG), the left superior temporal gyrus (left STG), and the right superior temporal gyrus (right STG). We used the same prior mean coordinates for the sources as in refs. 14 and 15 and likewise assumed extrinsic inputs (auditory stimuli) enter bilaterally into the primary auditory cortices. The models differed in the connections within and between the sources that were allowed to vary in strength with deviancy.

The network architecture for the first model tested was based on previous DCM studies of simple tone-deviancy effects (14, 15). Connections between HG and STG in both hemispheres and the intrinsic connectivity of primary auditory cortex were allowed to change with the deviant stimuli. This model did not allow for changes of intrinsic connections in the STG, a putative site for long-term representations of phoneme category. All of the remaining models allowed for intrinsic changes in STG. In model 2, modulation of the reciprocal connections between HG and STG and the intrinsic connectivity of STG were restricted to the left hemisphere. In model 3, these modulations were re-

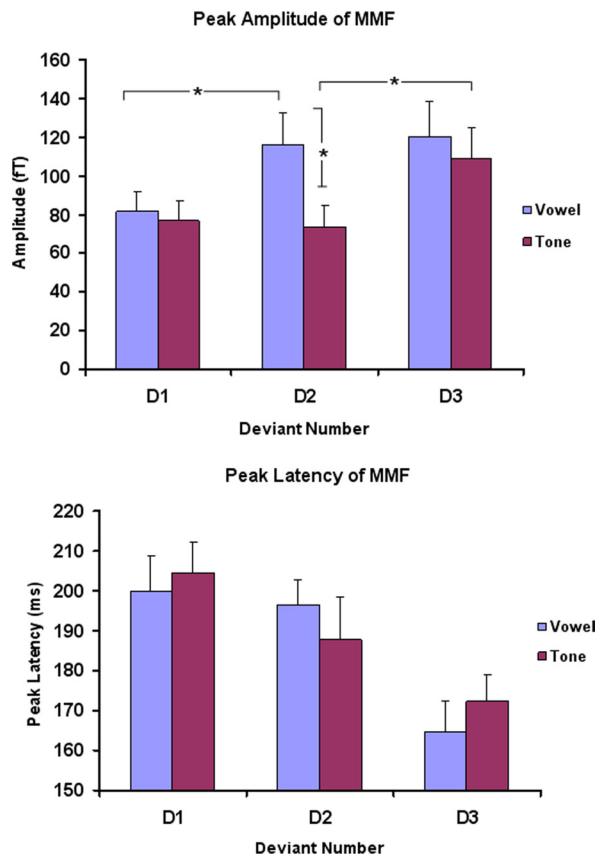


Fig. 2. MMF results for all 6 deviants compared with the relevant standard. (Upper) The amplitude responses (in femtoTesla) for the 3 deviants (D1, D2, and D3) across both stimulus classes (vowels in blue, tones in magenta). * denotes peak amplitude differences significant at $P < 0.05$. Note the relatively large amplitude response for vowel D2 compared with the perceptually matched tone D2. (Lower) The peak latency data. More distant deviants (across both classes) gave rise to quicker times to the MMF peak. These amplitude and latency summary statistics were taken from the maxima of spatiotemporal SPMs of F statistics, testing for a main effect of stimulus (D1, D2, and D3).

stricted to the right hemisphere. Model 4 was a fully bilateral model, in which all connection strengths in both hemispheres were allowed to change, including the interhemispheric connections between HG and STG. Interhemispheric connections were included because there is good evidence that primary auditory areas and caudal portions of the STG are connected to homotopic regions via the caudal corpus callosum (19). The 4 competing models, embodying different explanations for deviant-dependent difference in ERFs, were fitted to the data from all subjects and compared by using a Bayesian model selection procedure (20). Critically, this selection accounts for the relative complexity of the models (e.g., number of parameters) when comparing models.

Our main aim was to determine how neuronal interactions within the auditory hierarchy were influenced by deviancy (D2 and D3 vs. D1), under phonemic processing. We expected to see connectivity changes in higher levels (i.e., the superior temporal regions) that presumably learn phonemic regularity and furnish appropriate predictions for lower levels. We anticipated that the depth of hierarchical learning would be less for tone stimuli.

Results

Conventional Analyses of Evoked Responses in Sensor Space. Our preliminary analyses in sensor space confirmed our expectations.

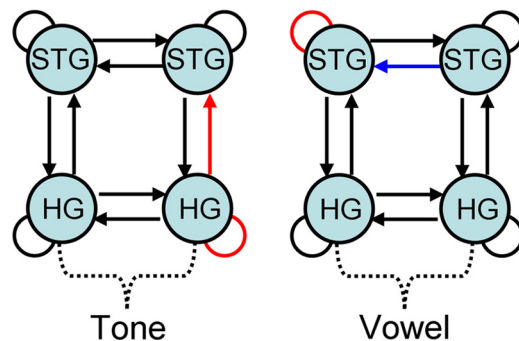


Fig. 3. Parameter estimates from the winning model (model 4) were used to test for deviancy-induced differences in connectivity within each stimulus class for tones (Left) and vowels (Right). Red indicates increasing connectivity, blue indicates decreasing connectivity.

We found that (i) all speech and nonspeech deviants produced mismatch responses; (ii) the peak amplitude of the mismatch increased with the degree of deviancy for both classes of stimuli, the amplitude for D3 being significantly greater than for D1 [vowels: $t(8) = 2.74$, $P < 0.05$; tones: $t(8) = 2.78$, $P < 0.05$]; and (iii) there was an effect of phonemic deviancy on the mismatch response for the vowel stimuli. In other words, there was an effect of stimulus class on the mismatch response for D2 deviants, the peak amplitude of the mismatch response to the vowel D2 being significantly greater than that elicited by the matched tone D2 [$t(8) = 2.64$, $P < 0.05$]. The 2 phonemic deviants (vowel D2 and D3) elicited mismatch responses of similar amplitude [$t(8) = -0.38$, $P = 0.71$] (see Fig. 2 and *SI Text* and Figs. S2 and S3 for more details).

DCM. All deviants, regardless of class, were best explained by the most complex bilateral model (model 4; see *SI Text* for details of the model comparison results). Note that this result is not simply because this model is the most complex of all models tested because the Bayesian model comparison approach we used takes model complexity into account. Within this bilateral model, the phonemic deviants (vowel D2 and D3) caused an increase in connectivity within the left STG and a decrease in the influence of the right STG upon the left STG (Fig. 3); these effects were stronger for D2 and D3 than for vowel D1 [within left STG: $t(8) = 2.56$, $P < 0.05$; right STG upon left STG: $t(8) = -2.36$, $P < 0.05$]. Within the tone class of stimuli, the most distant frequency deviant (D3) produced a mismatch response of greater amplitude than the other frequency deviants (D1 and D2), which was caused by an increase in connectivity within the right HG and an increase in the influence of HG on the right STG via forward connections [within right HG: $t(8) = 2.25$, $P = 0.05$; right HG upon right STG: $t(8) = 4.06$, $P < 0.05$] (see Fig. 3 and *SI Text* for more details).

Discussion

In this study, we investigated differences in neural responses to auditory stimuli that have high-level neural representations (vowels) with those that do not (tones). We used an oddball paradigm, which takes advantage of automatic change detection and implicit perceptual learning processes occurring within auditory cortex when a series of identical stimuli (standards) are interrupted unpredictably by a different stimulus (oddball or deviant).

When the MMF amplitudes for the 3 vowel deviants were tested against their tone counterparts, we found no effect of stimulus class for D1 and D3 (Fig. 2), with only D2 differing across classes. The larger relative response for this vowel deviant compared with the matched tone D2 deviant is likely because it

was perceived as a different phoneme to its standard. This finding agrees with previous studies that have shown that, for speech sounds, the amplitude of the mismatch response depends on both native phoneme category membership and absolute acoustic difference from the standard (21). In common with Vihla et al. (22), we found no significant difference in the mismatch response amplitude associated with the phoneme category deviants, D2 and D3, despite these being further away from each other in terms of acoustic difference than D2 is from the standard. Therefore, the main perceptual driver for mismatch response to D2 and D3 appears to be phoneme change rather than acoustic difference. In contrast, although the frequency difference of tone D2 from the standard was slightly greater than that of tone D1, the MMF amplitudes were not significantly different. The most distant tone frequency deviant, D3, had significantly greater MMF amplitude than either tone D1 or tone D2 (Fig. 2).

We then carried out a DCM analysis to explore how these mismatch responses could best be explained in terms of changes in the strength of the connectivity within and between the cortical sources of the auditory hierarchy. Our network consisted of 4 sources within the supratemporal plane: the primary auditory cortex and posterior STG in the left and right hemispheres (15, 16). We tested a model derived from the best previously published model of the neuronal dynamics of this system (14) against 3 models that allowed for deviant-induced variation in intrinsic connectivity within STG. These 3 models included left and right lateralized models and a more complex “fully connected” model that additionally allowed interhemispheric connections between both the HG and STG to vary in strength and thereby explain the mismatch response (Fig. 1). The more complex bilateral model provided a better explanation of the observed data than the simpler models.

The highest level in these models was auditory cortex. We did not include the right inferior frontal gyrus (IFG), commonly implicated in mismatch studies, because the prevailing view is that the temporal and frontal activations are functionally distinct and operate at different time scales (23, 24), with the temporal generators sensitive to stimulus change, and activity within the frontal generators reflecting a later shift of attention to the novel stimulus. MEG (unlike EEG) has been shown to be insensitive to the putative “attention shift” frontal sources (25), therefore 4 temporal sources were considered sufficient to model the data. To test this assumption, we also estimated 2 additional 5-region models (one with the right IFG as the fifth source, one with the left IFG as the fifth source). As expected, these performed very poorly compared with all 4-region models (see *SI Text* for details).

For the next stage of the analysis we looked within the “winning” model to see how the significant differences in the mismatch response within stimulus class, as observed in sensor space, could be explained in terms of differences in neuronal interactions.

The results for the tone comparison (D3 vs. D1/D2), which reflects an increase in frequency deviancy, showed that 2 connections were stronger for D3: the intrinsic connections of the HG and the forward connection between the HG and STG, both on the right (Fig. 3). It should be stressed that the other connections in the model are still important for producing the mismatch response; it is just that they are not more (or less) modulated by one deviant compared with the others. This result is consistent with predictive coding theory; all tones, D1, D2, and D3, violated the prediction (accrued through repeated exposure to standard stimuli) in the same way; they were of a different frequency from the expected stimulus, but D3 violated the prediction to a much greater extent. The increase in postsynaptic sensitivity (intrinsic connectivity), within the right HG observed for this deviant, may reflect the generation of a corresponding

error signal of greater magnitude, and the increase in the influence of the right HG over the right STG may reflect the forward influence of this signal on the higher cortical level. Such effects of frequency deviancy being apparent at the level of the HG are consistent with the findings of Ulanovsky and colleagues (8), who recorded directly from neurons in cat primary auditory cortex and found an increase in neuronal response to deviant stimuli, with some neurons displaying a hyperacuity in sensitivity to frequency change.

The DCM result for the vowel responses is very different. For the phonemic deviancy contrast (D2/D3 vs. D1) there was one connection that was modulated to a significantly greater extent during the mismatch response and one connection that became significantly weaker: we observed an increase in connection strength within the intrinsic connection on the left STG and a decrease in the influence of the right STG upon the left STG (Fig. 3). Considered in a predictive coding framework, vowel deviants D2 and D3 violated the prediction in a different manner than did vowel deviant D1; vowel D1 differed from the expected stimulus only in formant frequency, whereas vowels D2 and D3 were, perceptually, from different phoneme categories. The native phonemes have neuronal representations that will have been built up and maintained by years of experience. Therefore, it is likely that the adjustments to the predictive model needed to account for these representations would be of a different nature to those needed to account for the simple frequency violation caused by vowel D1. From this perspective, it is not surprising that the changes observed occur within a higher level of the network, i.e., STG, which, from a predictive coding perspective, would be involved in adjusting the predictive model. The increase in connectivity seen in the left STG and the conjoint decrease in the influence of the right STG on the left suggests a decoupling of these regions, perhaps because the higher level is in this case generating predictions about phonemic as opposed to spectrotemporal features. Such representations are thought to be used by higher-level multimodal left hemisphere networks to access word form and meaning (26–31).

It has been argued recently that left and right auditory cortex are differentially sensitive to the spectrotemporal attributes of auditory stimuli, with the left hemisphere exhibiting relatively higher temporal resolution and lower spectral resolution and the right hemisphere having a lower temporal resolution and a higher spectral resolution (32). According to this view, a left hemisphere bias for speech processing may arise from the high incidence of fine temporal information in the speech signal. Our findings challenge this hypothesis. At the spectrotemporal level, all our deviants differed from the standard only in the frequency dimension. Our tone stimuli did indeed cause an increase in coupling within the right hemisphere with increasing frequency deviancy as predicted by this hypothesis. However, the connectivity increase associated with our phoneme deviants occurred within the left hemisphere. Our results therefore suggest a left hemisphere bias for the processing of phonologically meaningful stimuli, regardless of their spectrotemporal characteristics.

Conclusion

We used a paradigm that exploits automatic change detection and rapid perceptual learning known to occur in auditory cortex. Our findings offer a perspective on how the areas in the auditory hierarchy exchange neuronal signals when meaningful auditory objects are processed. Meaningful changes in the speech signal evoke changes within higher levels of auditory cortex in the left hemisphere, whereas equivalent spectrotemporal changes in nonspeech stimuli induce connectivity changes within primary auditory cortex in the right hemisphere.

Materials and Methods

Subjects and Stimuli. Nine right-handed native English-speaking subjects participated (mean age 28, 3 male, 6 female). No subjects were hearing impaired or had any preexisting neurological or psychiatric disorder. All participants gave written consent, and the study was approved by the National Hospital for Neurology and Neurosurgery and Institute of Neurology joint research ethics committee.

Vowel Stimuli. The synthesized word stimuli took the form of [b/ + vowel +t/]. The stimuli were based on those from a previous study (33) and were designed to model a recording of a male British English speaker (the /b/ burst and the /t/ release were excised from a natural recording of this speaker rather than being synthesized). The duration of each word was 464 ms (260 ms for the vowel, excluding the bursts and /t/ stop gap). A total of 29 stimuli were produced, the vowel sound varying in F1 and F2 frequencies for each. The vowels were synthesized by using the cascade branch of a Klatt synthesizer (34). F0 had a falling contour from 152 to 119 Hz. The formant frequencies of F3–F5 were 2,500, 3,500, and 4,500 Hz. The bandwidths of the formants were 100, 180, 250, 300, and 550 for F1–F5. The frequencies of the first and second formants of the vowel sound in the standard stimulus (F1: 628 Hz, F2: 1,014 Hz) created a prototypic /a/ vowel [as judged in a previous study (33)]. The deviant stimuli differed from the standard (see Fig. S1) in a nonlinear, monotonic fashion. The formant frequencies of the vowel sounds in the deviants were chosen so that F1 and F2 fell along a vector through the equivalent rectangular bandwidth-transformed vowel space [ERB: a scale based on critical bandwidths in the auditory system, so differences are more linearly related to perception (35)], with the formant frequencies of the furthest deviant matching that of a prototypic /i/.

Perceptual thresholds were measured outside the scanner in a quiet room and before the subjects had been exposed to the experimental paradigm. Stimuli were delivered via an Axim X50v PocketPC computer using Sennheiser HD 650 headphones. The volume was adjusted to a comfortable level for each subject. Subjects completed a same-different discrimination task with a fixed standard; the deviant stimulus was varied adaptively (36) to find the perceptual threshold (i.e., acoustic difference where subjects correctly discriminated the stimuli on 71% of the trials). The mean perceptual threshold in terms of stimulus distance from the standard (in ERB) for vowel stimuli was 0.61 (SD 0.22), which was well below the distance between the standard and closest vowel deviant.

All of the stimuli were used when testing the subjects' perceptual thresholds behaviorally, but only 4 were used in the mismatch paradigm, the standard stimulus and 3 deviant stimuli. The Euclidean distances of the deviants from the standard were 1.16, 2.32, and 9.30 ERB, respectively. Qualitatively, this meant that the first deviant (D1) was above all subjects' behavioural discrimination thresholds, but still within the same vowel category, the second deviant (D2) was twice as far away from the standard as D1 and similar to the distance that would make a categorical difference between vowels in English (33), and the third deviant (D3) represented a stimulus on the other side of the vowel space (i.e., 8 times further away from the standard than D1, as large a difference in F1 and F2 as it is possible to synthesize). The first and second formant frequencies of the 3 deviants were, respectively, 565, 507, and 237 Hz for F1, and 1,144, 1,287, and 2,522 Hz for F2.

The identity of these 4 stimuli were checked in a separate 4-way, forced-choice experiment on a different set of subjects ($n = 11$) who were native English speakers. This procedure confirmed that the standard was perceived as Bart (100%); D1 was also perceived as Bart (87%); D2 was perceived as Burt (100%); and D3 was perceived as Beat (100%).

Tone Stimuli. Frequency discrimination is better for sinusoids than for vowel formants of the same frequency because sinusoids are narrow-band whereas formants are broad-band. For formant shifts there is no change in the carrier frequency but for tones there is; tone changes are thus easier to detect because they involve changes in pitch and frequency shifts in the spectral envelope (e.g., see ref. 37). The tone stimuli used in this study were sinusoids of 234-ms duration that were amplitude modulated by one-half cycle of a raised cosine (i.e., leading to a gradually rising then falling amplitude envelope). The frequency of the standard stimulus was the same as for the F2 of the vowel standards (1,014 Hz). To correctly discriminate stimuli on 71% of trials, subjects required, on average, a 23-Hz difference in the tone stimuli, but a simultaneous change of 33 Hz in F1 and 66 Hz in F2 for the vowel. Our tone deviant stimuli were chosen so that they would shadow the vowel stimuli in terms of relative distances from threshold. That is, even though the tone frequency differences were smaller than for the vowels, as with the vowel stimuli, D1 (1,044 Hz) was slightly above the behavioral threshold, D2 (1,074

Hz) was twice as far away from the standard as D1, and D3 (1,278 Hz) was 8 times as far away from the standard as D1.

MEG Scanning and Stimulus Presentation. A VSM MedTech Omega 275 MEG scanner was used to measure the electromagnetic field changes that occurred during the experiment from 275 superconducting quantum interference devices (each referenced to a third-order axial gradiometer) arranged around the head. A total of 480 data points were sampled each second with an antialias filter applied at 120 Hz.

Auditory stimuli were presented binaurally by using E-A-RTONE 3A audiometric insert earphones (Etymotic Research) that were attached to the rear of the subject's chair and connected to the subject with flexible tubing. The stimuli were presented initially at 60 dB/sound pressure level. Subjects were allowed to alter this to a comfortable level while listening to the stimuli during a test period. A passive odd-ball paradigm was used involving the auditory presentation (stimulus onset asynchrony = 1,080 ms) of a train of repeating standards interleaved in a pseudorandomized manner with presentations of D1, D2, or D3. Within each acquisition block 30 deviants (of each sort) were presented to create a standards-to-deviants ratio of 4:1. A minimum of 2 standards was presented between deviants. A total of 6 acquisition blocks were performed by each subject, 3 blocks with vowel stimuli, 3 with tone stimuli, resulting in a total of 90 trials for each of the 6 deviants. Each block lasted for 540 s.

During stimulus presentation subjects were asked to complete an incidental visual detection task and not to pay attention to the auditory stimuli. Static pictures of outdoor scenes were presented for 60 s followed by a picture (presented for 1.5 s) of either a circle or a square (red shape on a gray background). The subjects were asked to press a response button (right index finger) for the circles (90%) and to withhold the response when presented with squares (10%). This go/no-go task provided evidence that subjects were attending to the visual modality (mean accuracy = 96%).

Statistical Analyses in Sensor Space. Statistical parametric mapping was performed by using SPM8b software (Wellcome Trust Centre for Neuroimaging), running under Matlab 7.4.0 (Mathworks). For each subject, the electromagnetic field data from all channels for each acquisition block were digitally filtered with a third-order butterworth band-pass of 1–20 Hz. Data were then organized into epochs referenced to stimulus presentation, running from 100 ms prestimulus to 600 ms poststimulus. The mean amplitude of the prestimulus period was used as a baseline for the poststimulus data points. Epochs that contained a value $\geq \pm 3,000$ femto-Tesla were rejected as artifacts and removed from further analysis. MEG data for each subject were inspected visually before averaging to ensure that the remaining trials did not contain eye blinks or eye movements that would materially affect the data upon averaging.

A 3D image was constructed for each epoch, electromagnetic field values being interpolated between sensors across 2D sensor space at each sample point in time. The data at each sensor at each time point were projected onto a 2D scalp map. This 2D scalp map was a 64×64 grid. In other words, there were 64 bins from the leftmost sensor to the rightmost sensor and 64 bins from the most anterior sensor to the most posterior sensor. These spatiotemporal images were entered into statistical parametric mapping analysis, for each subject, with 4 levels of 1 factor (i.e., STD, D1, D2, D3). The search space included all voxels in sensor space between the time points 150–250 ms poststimulus presentation (18). The data were interrogated with F-contrasts of the form $[-1 \ 1 \ 0 \ 0]$ to identify the MMF for D1, $[-1 \ 0 \ 1 \ 0]$ to identify the MMF for D2, and $[-1 \ 0 \ 0 \ 1]$ to identify the MMF for D3. The value of the contrast estimate at the peak voxel in the resulting F-maps, and the peak latency at this voxel was extracted for each contrast to provide summary statistics for each deviant (i.e., contrast) and subject. This process was repeated for both vowel and tone stimuli.

To investigate the effects of deviancy within class, the amplitude summary statistics for each deviant level were tested against each other by using paired t tests across subjects. Peak amplitudes for the vowel stimuli were tested against the equivalent tone estimates with paired t tests (across subjects) to look for effects of stimulus class at each level of deviancy. The peak latencies for the vowel stimuli were also tested against the peak latencies for the equivalent tone stimuli with paired t tests across subjects.

DCM. DCM estimates the influence of cortical subpopulations of neurons on one another by the inversion of a biologically informed, spatiotemporal forward model of the observed MEG sensor-space activity. The concept of applying a spatiotemporal dipole model was first put forward by Scherg (38). In DCM, the parameterization of the temporal model is based on neural mass models, the details of which are explained at length in refs. 12, 39, and 40.

Simply put, DCM attempts to explain the distribution and timing of the field changes observed in sensor space, after stimulus presentations, by combining a user-specified network architecture (i.e.: the number and location of the sources in the network, how these sources are connected, where the inputs enter the system) with empirically derived information about neuronal dynamics.

For each subject, the average ERF for each deviant was calculated from the epoched data used for the sensor space analysis (Fig. S2). The data from each of these ERFs, between stimulus presentation and 300 ms poststimulus, were reduced to 16 spatial modes by singular value decomposition and used to invert each dynamic causal model. Each subject-specific DCM modeled one deviant, and the difference between deviants was explained in terms of changes in coupling specified by the particular model. Using the procedure specified by Friston et al. (41), the log evidence was evaluated for all models

for each deviant class (vowel and tone) and used to compare models. The ensuing differences in log evidence (i.e., log Bayes factors or marginal likelihood ratios) were summed across subjects (given the data were acquired independently from each subject) to pool log evidence over subjects. A relative log evidence >3 was taken as strong evidence in favor of one model over another (20). Once the best model was identified, the deviant-dependent coupling change was extracted for each connection and each type of stimulus. These parameter estimates were used as subject-specific summary statistics as above and fed into a second-level analysis, testing for consistent differences under vowels and tones, across subjects, by means of paired *t* tests.

ACKNOWLEDGMENTS. This work was supported by The Wellcome Trust and National Institute for Health Research Comprehensive Biomedical Research Centre at University College Hospitals and the University Research Priority Program Foundations of Social Human Behavior at the University of Zurich.

1. Clark J, Yallop C, Fletcher J (2007) *An Introduction to Phonetics and Phonology* (Blackwell, Oxford).
2. Näätänen R, Gaillard AW, Mantysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)* 42:313–329.
3. Näätänen R, Jacobsen T, Winkler I (2005) Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology* 42:25–32.
4. Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
5. Opitz B, Mecklinger A, Von Cramon DY, Kruggel F (1999) Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology* 36:142–147.
6. Opitz B, et al. (2002) Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage* 15:167–174.
7. Javitt DC, Steinschneider M, Schroeder CE, Arezzo JC (1996) Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia. *Proc Natl Acad Sci USA* 93:11962–11967.
8. Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391–398.
9. Friston K (2005) A theory of cortical responses. *Philos Trans R Soc London Ser B* 360:815–836.
10. Savelle J, et al. (2003) The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Brain Res Cognit Brain Res* 16:250–256.
11. Winkler I, et al. (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36:638–642.
12. Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage* 30:1273–1284.
13. Kewley-Port D, Watson CS (1994) Formant-frequency discrimination for isolated English vowels. *J Acoust Soc Am* 95:485–496.
14. Garrido MI, et al. (2008) The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage* 42:936–944.
15. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci USA* 104:20961–20966.
16. Garrido MI, et al. (2007) Dynamic causal modeling of evoked potentials: A reproducibility study. *NeuroImage* 36:571–580.
17. Kiebel SJ, Garrido MI, Friston KJ (2007) Dynamic causal modeling of evoked responses: The role of intrinsic connections. *NeuroImage* 36:332–345.
18. Kiebel SJ, Kilner J, Friston KJ (2007) in *Statistical Parametric Mapping*, eds Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny, WD (Academic, London), pp 211–231.
19. Cipolloni PB, Pandya DN (1985) Topography and trajectories of commissural fibers of the superior temporal region in the rhesus monkey. *Exp Brain Res* 57:381–389.
20. Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. *NeuroImage* 22:1157–1172.
21. Näätänen R (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38:1–21.
22. Vihla M, Lounasmaa OV, Salmelin R (2000) Cortical processing of change detection: Dissociation between natural vowels and two-frequency complex tones. *Proc Natl Acad Sci USA* 97:10590–10594.
23. Deouell LY (2007) The frontal generator of the mismatch negativity revisited. *J Psychophysiol* 21:188–203.
24. Rinne T, et al. (2000) Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroImage* 12:14–19.
25. Hamalainen M, et al. (1993) Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65:413–497.
26. Spitsyna G, et al. (2006) Converging language streams in the human temporal lobe. *J Neurosci* 26:7328–7336.
27. Vigneau M, et al. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* 30:1414–1432.
28. Binder JR, et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
29. 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.
30. Uppenkamp S, et al. (2006) Locating the initial stages of speech-sound processing in human temporal cortex. *NeuroImage* 31:1284–1296.
31. Leff AP, et al. (2008) The cortical dynamics of intelligible speech. *J Neurosci* 28:13209–13215.
32. Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and speech. *Trends Cognit Sci* 6:37–46.
33. Iverson P, Evans BG (2007) Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *J Acoust Soc Am* 122:2842–2854.
34. Klatt DH, Klatt LC (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am* 87:820–857.
35. Glasberg BR, Moore BC (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47:103–138.
36. Levitt H (1971) Transformed up-down methods in psychoacoustics *J Acoust Soc Am* 49(Suppl 2):467.
37. Plack CJ, Oxenham AJ, Fay RR, Popper AN (2005) *Pitch: Neural Coding and Perception* (Springer, New York).
38. Scherg M, Von Cramon D (1985) Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol* 62:32–44.
39. David O, Friston KJ (2003) A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage* 20:1743–1755.
40. Fastenrath M, Friston KJ, Kiebel SJ (2009) Dynamical causal modeling for M/EEG: Spatial and temporal symmetry constraints. *NeuroImage* 44:154–163.
41. Friston K, et al. (2007) Variational free energy and the Laplace approximation. *NeuroImage* 34:220–234.