

Dynamic Causal Modeling of the Response to Frequency Deviants

Marta I. Garrido, James M. Kilner, Stefan J. Kiebel, and Karl J. Friston

Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom

Submitted 21 February 2008; accepted in final form 24 February 2009

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol* 101: 2620–2631, 2009. First published March 4, 2009; doi:10.1152/jn.90291.2008. This article describes the use of dynamic causal modeling to test hypotheses about the genesis of evoked responses. Specifically, we consider the mismatch negativity (MMN), a well-characterized response to deviant sounds and one of the most widely studied evoked responses. There have been several mechanistic accounts of how the MMN might arise. It has been suggested that the MMN results from a comparison between sensory input and a memory trace of previous input, although others have argued that local adaptation, due to stimulus repetition, is sufficient to explain the MMN. Thus the precise mechanisms underlying the generation of the MMN remain unclear. This study tests some biologically plausible spatiotemporal dipole models that rest on changes in extrinsic top-down connections (that enable comparison) and intrinsic changes (that model adaptation). Dynamic causal modeling suggested that responses to deviants are best explained by changes in effective connectivity both within and between cortical sources in a hierarchical network of distributed sources. Our model comparison suggests that both adaptation and memory comparison operate in concert to produce the early (N1 enhancement) and late (MMN) parts of the response to frequency deviants. We consider these mechanisms in the light of predictive coding and hierarchical inference in the brain.

INTRODUCTION

This study attempts to compare alternative hypotheses about the mismatch negativity (MMN) with dynamic causal modeling (DCM), a method for the analysis of effective connectivity in the brain (David et al. 2006; Friston et al. 2003). DCM is a hypothesis-driven approach that allows one to infer the causal architecture of directed neuronal interactions. We considered several dynamic causal models (DCMs), representing distinct hypotheses about the generation of electromagnetic signals in an MMN paradigm and quantified their relative evidence (Friston et al. 2006; Penny et al. 2004). DCMs are spatiotemporal dipole models (Scherg and Van Cramon 1985) that use a conventional formulation of source activity (cf. equivalent current dipole [ECD] models; Kiebel et al. 2006), but place constraints on the way signals are generated. These constraints require activity in one source to be caused by activity in another (David et al. 2006), where the implicit neuronal interactions rest on neural mass models. Having previously established the face validity of DCM for event-related potentials (ERPs) (David et al. 2006; Kiebel et al. 2006) and the predictive validity—in terms of reproducibility across subjects (Garrido et al. 2007a)—we show here how DCM and Bayesian model comparison can be used to adjudicate among different models

of evoked responses and, in particular, responses to frequency deviants.

Mismatch negativity

Deviants (oddballs) embedded in a stream of repeated events (standards) produce distinct responses that can be recorded noninvasively with electrophysiological techniques such as electroencephalography (EEG) and magnetoencephalography (MEG). The mismatch negativity (MMN) is the negative component of the waveform obtained by subtracting the event-related response to a standard from the response to an oddball. This response to changes in stimulus features peaks at about 100–200 ms from change onset (Sams et al. 1985) and has a negativity that is distributed over frontocentral scalp areas, with prominence in frontal regions.

The MMN is believed to be an index of automatic change detection, reflecting a preattentive sensory memory mechanism (Tiitinen et al. 1994). There have been several mechanistic accounts of how the MMN might arise. Early work by Näätänen and colleagues suggested that the MMN results from a comparison between auditory input and a memory trace of previous sounds. They postulated that the MMN reflects an on-line modification or update of a perceptual model, on encountering a deviant stimulus—the *model-adjustment hypothesis* (Näätänen and Winkler 1999; Winkler et al. 1996). Thus the MMN is a specific response to stimulus change and not to the stimulus per se. Recently, Jääskeläinen et al. (2004) challenged this view by suggesting that responses to frequency deviants could be mediated by a much simpler mechanism—that is, release from local adaptation confined to the auditory cortex, where repetition of standard tones attenuates the N1 response. The enhanced N1 response to an oddball could therefore be misinterpreted as a component that was distinct from the N1. This is the *adaptation hypothesis*. According to this view, neuronal populations in the auditory cortex become less responsive with repeated stimulation, which is sufficient to explain the MMN. In short, standards suppress the N1 through adaptation in the auditory cortex and the MMN emerges as a subtraction artifact (see also Rosburg et al. 2004).

More recently, we have proposed that the MMN might be understood in terms of perceptual learning under hierarchical generative models of auditory input (Baldeweg 2006; Friston 2005). This explanation rests on a predictive coding scheme for perceptual inference and learning that entails suppressing prediction error using top-down predictions. Perceptual learning, during the repeated presentation of standards, involves changes in both intrinsic (within-area) and extrinsic (between-area) connections that make the “explaining away” of prediction error more efficient. Critically, this explanation rests on key elements of both the model-adjustment and adaptation hypotheses: Under predictive coding, the prediction error depends on

Address for reprint requests and other correspondence: M. I. Garrido, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, UK WC1N 3BG (E-mail: m.garrido@fil.ion.ucl.ac.uk).

a comparison between the current stimulus and a prediction based on a memory or model of previous standards (encoded by changes in extrinsic connectivity). It is this prediction error that causes the MMN and subsequent model adjustment by inducing associative plasticity. In predictive coding, synaptic connection strengths change to minimize prediction error over repeated exposure to stimuli. This involves changes in both extrinsic and intrinsic connections that model the causes and dynamics of sensory input, respectively (see Friston 2008 for details). In addition to these changes, the postsynaptic sensitivity or gain of units encoding prediction error are optimized to reflect the precision of that error. Efficiently, this involves adaption of error units such that prediction error is attenuated when it has a high variance or low precision. The combined effects of changes in synaptic efficacy and sensitivity suppress prediction error over repeated exposure to standards (MI Garrido, JM Kilner, SJ Kiebel, KE Stephan, T Baldeweg, and KJ Friston, unpublished observations), resulting in an exuberant prediction error signal when a deviant is encountered, which is thought to be the basis of the MMN. In short, predictive coding calls on both comparison and adaptation.

Predictive coding introduces the notion of a hierarchy and the distinction between bottom-up afferents and top-down predictions. This is not an integral part of the model-adjustment hypothesis; however, there is some evidence to suggest the MMN relies on hierarchically distributed sources, possibly extending into the prefrontal cortex. For example, Giard et al. (1990) discussed a topographic study of auditory deviance in light of Näätänen's model and suggested that the MMN is caused by two underlying functional processes: a sensory memory mechanism related to temporal generators and an automatic attention-switching process related to frontal generators. It has also been shown that the temporal and frontal MMN sources have distinct behaviors over time (Rinne et al. 2000) and that these sources interact with each other (Jemel et al. 2002). Thus responses to change could be generated by a temporofrontal network as revealed by EEG and functional magnetic resonance imaging (MRI) studies (Doeller et al. 2003; Giard et al. 1990; Molholm et al. 2005; Optiz et al. 2002; Rinne et al. 2000, 2005; Tse and Penney 2008).

It should be noted that the key argument in the literature is between Näätänen's group—who posit that the MMN proper is elicited by a mechanism (and neurons) distinct from that which elicits the N1—and those claiming there is only one mechanism that generates both components. This distinction does not speak to the presence or absence of top-down predictions proposed by predictive coding; indeed the model-adjustment hypothesis could be mediated by changes in the efficacy of connections intrinsic to auditory sources (e.g., Näätänen and Winkler 1999). For example, Maess et al. (2007) showed that the sensory (adaptation) part of responses to deviant frequencies is maximal in the late N1 range (105–125 ms), whereas the comparison part peaks in the late MMN range (170–200 ms). Inverse modeling revealed sources for both contributions in bilateral temporal gyri. However, in what follows, we will assume that the model (in the model-adjustment hypothesis) is furnished by top-down predictions and therefore depends on extrinsic connections.

Overview of the present study

Here, we test competing mechanistic hypotheses, or models, framed in terms of repetition-dependent changes in coupling (i.e., plasticity; cf. May et al. 1999) and used Bayesian model comparison of DCMs to find the best model. The models we examined explained responses to deviants in terms of 1) changes in postsynaptic sensitivity intrinsic to the auditory cortex (cf. adaptation; Jääskeläinen et al. 2004); 2) changes in extrinsic connections entailed by model-adjustment (cf. Doeller et al. 2003; Winkler et al. 1996); and 3) a combination of both, as suggested by predictive coding (cf. Baldeweg 2006; Friston 2005). Our results suggest that rather than being mutually exclusive, both intrinsic (adaptation) and extrinsic (model-adjustment) changes in connectivity are required to explain the data observed. These results are discussed in terms of predictive coding and hierarchical inference in the brain.

METHODS

Modeling and model comparison

Predictive coding encompasses adaptation (Jääskeläinen et al. 2004; May et al. 1999) and model-adjustment (Näätänen and Winkler 1999; Winkler et al. 1996). Although the latter allows for adaptation effects (which the authors refer to as refractoriness), the adaptation hypothesis precludes a prediction or model-dependent contribution to the MMN. Predictive coding entails both perceptual learning or adjustment (to encode predictions) and adaptation (to encode the increasing precision of these predictions). We tested the relative contributions of model adjustment and adaptation by formulating them as network models with plastic changes in extrinsic (model-adjustment) and intrinsic (adaptation) connections. We hoped to show that both contribute to responses to deviants, consistent with predictive coding or model-based explanations (Baldeweg 2006; Friston 2005; Winkler 2007).

DYNAMIC CAUSAL MODELING. DCM is an important advance over conventional analyses of evoked responses because it places natural constraints on the inversion; that is, activity in one source has to be caused by activity in another (see David and Friston 2003; David et al. 2005). DCMs for EEG and MEG use neural mass models to explain source activity. In other words, DCM does not simply estimate dipole activity at a particular point in time; it tries to model dipole activity over time, in terms of interacting inhibitory and excitatory populations of neurons. The dynamics of these populations are based on the model of Jansen and Rit (1995). We use a neural mass model with extrinsic connections among distributed sources (David et al. 2005) that conforms to the connectivity rules described in Felleman and Van Essen (1991). These rules allow one to build networks of coupled sources, where each source comprises three neuronal populations or masses. Condition-specific effects are modeled in terms of changes in coupling between or within sources. Changes in connections among populations, within a source, allow one to model local adaptation. This is modeled by changes in a synaptic parameter, which controls the peak amplitude of postsynaptic responses to a population's excitatory inputs (see Kiebel et al. 2007 for details).

A DCM is specified in terms of some state equations that summarize the average synaptic dynamics in terms of spike-rate-dependent current and voltage (i.e., state) changes, for all populations in all sources

$$\dot{x} = f(x, u, \theta) \quad (1)$$

This means that the evolution of neuronal states x is a function (parameterized by θ) of the current electrophysiological state and exogenous (i.e., subcortical) input u . The input usually encodes the

presentation of a stimulus and models thalamic input to specific cortical sources. In this work, we used a Gaussian bump function, whose onset and dispersion were free parameters (for details see David and Friston 2003; David et al. 2005, 2006). The state equations embody the connection rules described earlier, where θ includes synaptic and connectivity parameters and their condition or stimulus-specific changes. Finally, an output equation couples specific states (the average depolarization of pyramidal cells in each source) x_0 , to the EEG signals y using a conventional linear electromagnetic forward model

$$y = L(\theta)x_0 + \varepsilon \quad (2)$$

This equation links the neuronal states to observed channel data. In this application, the lead field $L(\theta)$ was parameterized in terms of the location and orientation of each source as described in Kiebel et al. (2006).

BAYESIAN MODEL COMPARISON. Inversion or fitting of a specific DCM, m , involves approximating the posterior probability on the parameters, which is proportional to the probability of the data (the likelihood), times the prior probability on the parameters

$$p(\theta|y, m) \propto p(y|\theta, m)p(\theta|m) \quad (3)$$

This approximation uses variational Bayes as described in Friston (2002) and can be formulated as a coordinate descent on free energy F . The aim is to minimize $F \approx -\ln p(y|m)$, which is an upper bound on the negative log-evidence. The best model, given the data, is the one with highest log-evidence, $\ln p(y|m)$. One can compare two models m_1 and m_2 using their Bayes factor (Penny et al. 2004) or, equivalently, the relative log-evidence: $\ln p(y|m_1) - \ln p(y|m_2)$. If this difference is >3 (i.e., their relative likelihood is $>20:1$) then one asserts there is strong evidence in favor of the first model. Note that the free energy takes into account both the accuracy and the complexity of models, favoring the first and penalizing the latter. Therefore one can compare models with different degrees of complexity or number of parameters (Friston et al. 2006). To select the model that best explains multiple data sets (i.e., the best model for data from several subjects) one simply multiplies the marginal likelihoods or, equivalently, adds the log-evidences from each subject

$$\ln p(y_1, \dots, y_n, m_i) = \sum_{j=1}^n \ln p(y_j|m_i) \quad (4)$$

This gives the log-evidence for the i th model across all n subjects (Garrido et al. 2007), assuming that all the subjects generated data under the same model.

RELATIONSHIP TO CONVENTIONAL ANALYSES. DCM operates on the channel data in the same way as sources are reconstructed in a conventional setting. In both cases, one needs to specify the number of sources and how their activity is expressed in sensor space. The only difference is that, with DCM, one also has to specify the connections among the sources and which sources receive exogenous input. In fact, inverting a DCM subsumes the inversion of a conventional forward model. This is because DCM has two components: a neural-mass model of the interactions among ECD sources and a classical electromagnetic forward model that links these sources to sensor measurements. Inverting the DCM implicitly optimizes the location and orientation of the sources. Indeed, if we removed the neuronal constraints on the model, DCM would reduce to a spatio-temporal dipole model (Scherg and Van Cramon 1985). Furthermore, using the aforementioned log-evidence, one can optimize the number of sources by comparing DCMs with different architectures; we will see an example of this in the following text.

Experimental design and analysis

SUBJECTS. We studied a group of 12 right-handed healthy volunteers aged 24–34 yr (4 female). Each subject gave signed informed consent before the study, which proceeded under local ethical committee guidelines.

TASK. Subjects sat on a comfortable chair in front of a desk in a dimly illuminated room. Electroencephalographic activity was measured during an auditory oddball paradigm, in which subjects heard of standard (500 Hz) and deviant tones (550 Hz), occurring 80% (480 trials) and 20% (120 trials) of the time, respectively, in a pseudorandom sequence (subject to the constraint that two deviant tones did not occur together). The stimuli were presented binaurally via headphones for 10 min every 0.5 s. The duration of each tone was 70 ms with 5-ms rise and fall times. Sound intensity was adjusted to a comfortable level for each subject. The subjects performed an incidental visual task and were instructed to ignore the sounds. The incidental task consisted of button pressing whenever the luminance contrast of a fixation cross changed (every 2 to 5 s). The visual changes did not coincide with auditory changes.

DATA ACQUISITION AND PROCESSING. EEG data were recorded with a *Biosemi* system at 128 scalp electrodes and a sampling rate of 512 Hz. Vertical and horizontal eye movements were monitored using electro-oculogram electrodes. The data were epoched off-line, with a peristimulus window of -100 to 400 ms, down-sampled to 200 Hz, band-pass filtered between 0.5 to 40 Hz, and rereferenced to an electrode attached to the tip of the nose. Artifact removal was implemented with robust averaging, which is a standard iterative scheme that produces the best estimate of the average by weighting data points as a function of their distance from the sample mean (cf. Wager et al. 2005). See Fig. 1A for a summary of these data averaged over subjects. For computational expediency, the dimensionality of the data were reduced to eight channel mixtures or spatial modes. These were the principal modes of a singular value decomposition of the channel data between 0 and 250 ms, from both trial types. The use of eight principal eigenvariates explained $>77\%$ of the data in all subjects and preserved the interesting components of evoked responses.

DCM SPECIFICATION. DCM is not an exploratory technique (i.e., it does not consider all possible models). DCM tests a family of specific models and, through Bayesian model selection, provides evidence in favor of one model, relative to others. The results of DCM depend on the models evaluated. This means there may be other equally plausible models, with different architectures (areas and connections). The networks we tested were motivated by previous studies of the MMN (Doeller et al. 2003; Giard et al. 1990; Molholm et al. 2005; Optiz et al. 2002; Rinne et al. 2000, 2005; Tse and Penney 2008). These suggest bilateral sources in the auditory cortex (A1), superior temporal gyrus (STG), and inferior frontal gyrus (IFG), which are usually stronger and identified more consistently in the right hemisphere.

We specified eight models, starting with a parsimonious model and adding hierarchical levels (i.e., sources and extrinsic connections). The simplest model comprised two sources or nodes in the left and right primary auditory cortices. We then added sources and connections, until we reached a symmetric three-level hierarchical model. All models can therefore be considered as special cases of this three-level hierarchical model.

These models attempt to explain condition-specific responses (i.e., responses to standards and deviants) under the constraint that differences among conditions have to be explained by, and only by, differences in coupling in specified connections. Therefore we allowed condition-specific differences (i.e., plasticity) in all extrinsic (between-source) connections. In addition, we considered every model with, and without, changes in intrinsic connections in bilateral A1. These sources were chosen as cortical input stations for process-

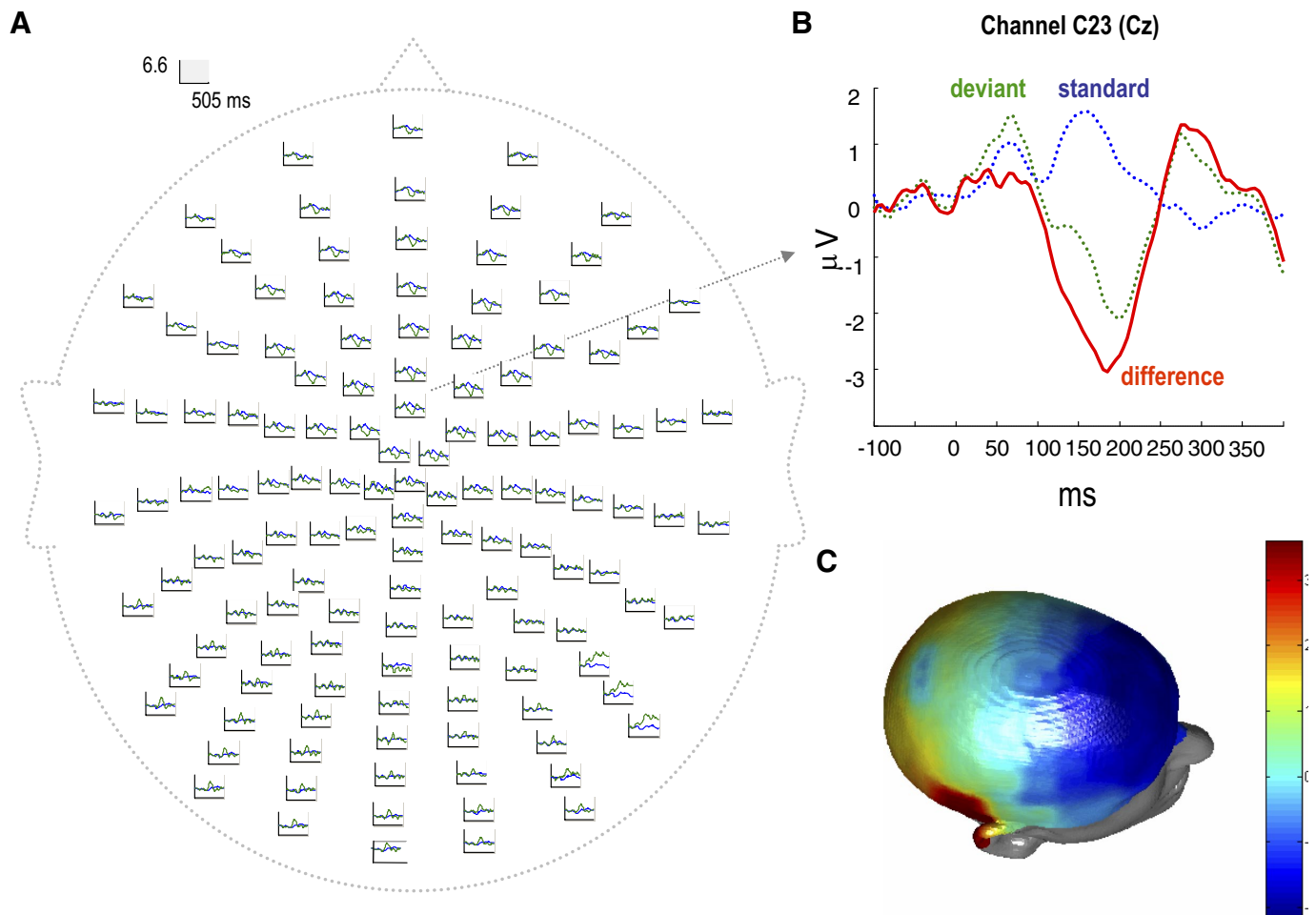


FIG. 1. Grand-mean event-related potentials (ERPs, averaged over subjects). *A*: ERP to the standard (blue) and deviant (green) tones overlaid on a whole scalp map of 128 electroencephalographic electrodes. *B*: ERP responses to the standard and deviant tones at electrode C23 (Cz). The difference waveform (red) was obtained by subtracting the grand-average ERP to standards from the ERP to deviants. *C*: grand-mean differences at peak (150 ms), interpolated to give a 3dimensional scalp topography.

ing auditory information. We modeled each source with a single equivalent current dipole (ECD) that corresponded to the activity of the pyramidal cell population in our neural mass model. Afferent auditory input was parameterized as a mixture of a Gaussian bump function (with unknown latency and dispersion). We modeled the data from 0 to 250 ms over peristimulus time as the response of each network to this input. The spatial part of the model used a four concentric sphere-head model with homogeneous and isotropic conductivity as an approximation to the brain, cerebrospinal fluid (CSF), skull, and scalp surfaces. The lead field, $L(\theta)$, which accounts for the contribution of cortical sources to measured signals, was parameterized in terms of the location and orientation of each source (Kiebel et al. 2006). This used the electromagnetic forward model solutions in the fieldtrip software (<http://www2.ru.nl/fcdonders/fieldtrip>). The coordinates reported by Opitz et al. (2002) and Rademacher et al. (2001) were chosen as prior source location means, with a prior variance of 16 mm^2 . We converted these coordinates, given in the literature in Talairach space, to Montreal Neurological Institute (MNI) space using the algorithm described in <http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach>. The moment parameters had a prior mean of zero and a variance of 256 mm^2 in each direction. This is equivalent to assuming uninformative or flat priors on the orientations of the dipole moments.

Our simplest model, S2, is a two-source network that corresponds to the hypothesis that ERPs to standards and deviants can both be explained by bilateral responses in primary auditory cortex (A1). This

model does not allow changes in connectivity and cannot model any responses to deviants. Model S2i is similar to S2 but allows for postsynaptic sensitivity changes intrinsic to A1. Here, we hypothesized that differences between responses to standards and deviants are caused by adaptation. Model S4 is a two-level hierarchical model comprising four sources. It is built on S2, to which we added left and right superior temporal gyri (STG) sources that were reciprocally connected (i.e., connected through forward and backward connections) to ipsilateral A1. Model S4i is the same as S4 but with plastic connections within A1. A three-level hierarchical model, comprising five sources (model S5), includes right inferior frontal gyrus (IFG). Right STG was reciprocally connected with ipsilateral IFG. Model S5i corresponds to S5 but has plastic connections within A1. Models S6 and S6i are extensions of S5 and S5i, respectively, which include the left IFG connected reciprocally to ipsilateral STG.

In summary, the models differed in terms of their nodes and in the connections, which could show putative repetition or learning-related changes. Models S2, S4, S5, and S6 allowed changes in extrinsic connections (apart from S2, which has no changeable connections and can be regarded as a null model). These models represent the hypotheses that ERP differences are due to changes in extrinsic connections that encode the expected auditory stimulus. Models S2i, S4i, S5i, and S6i allowed for changes in the extrinsic connections, plus changes in intrinsic connectivity or adaptation within left and right A1. These correspond to the hypotheses that differences in ERPs are due to conjoint changes in extrinsic and intrinsic connections.

Model S2 is a null model in which there are neither local synaptic changes (within an area) nor cortical interactions (between areas) (i.e., there is no MMN). See Fig. 2 for a schematic summary of these models.

RESULTS

Event-related potentials

Figure 1A shows the grand-mean responses (i.e., averaged across subjects) to the standard (probability of occurrence = 0.8; 500 Hz) and deviant tones (probability of occurrence = 0.2; 550 Hz). Responses to deviants exhibited positivities (peaking at about 50 and 250 ms) and a widespread negativity (peaking at 200 ms) over frontal and central electrodes. The MMN difference between the ERPs evoked by the standard and deviant tones had a typical scalp distribution (Fig. 1C) and peaked at 180 ms (Fig. 1B). These spatiotemporal responses over sensors are characteristic of MMN studies; the aim of this

study was to identify the cortical network that might generate these signals.

Bayesian model selection

The eight network models (DCMs) were inverted for each of the 12 subjects. Figure 3 shows the log-evidence across the eight models considered; this measures how good one model is compared with another. The model with the highest evidence explains the data with the best balance of accuracy and complexity. Our results show that model S5i (highlighted) is the model that best explains the group data. We found very strong evidence (a difference in log-evidence of >5) for this model, relative to the remaining models. The two models with a single hierarchical level (S2 and S2i) performed very poorly, compared with the hierarchical models. Model S5i was the best among the hierarchical models. This is an asymmetrical three-level hierarchical network, comprising five extrinsically inter-

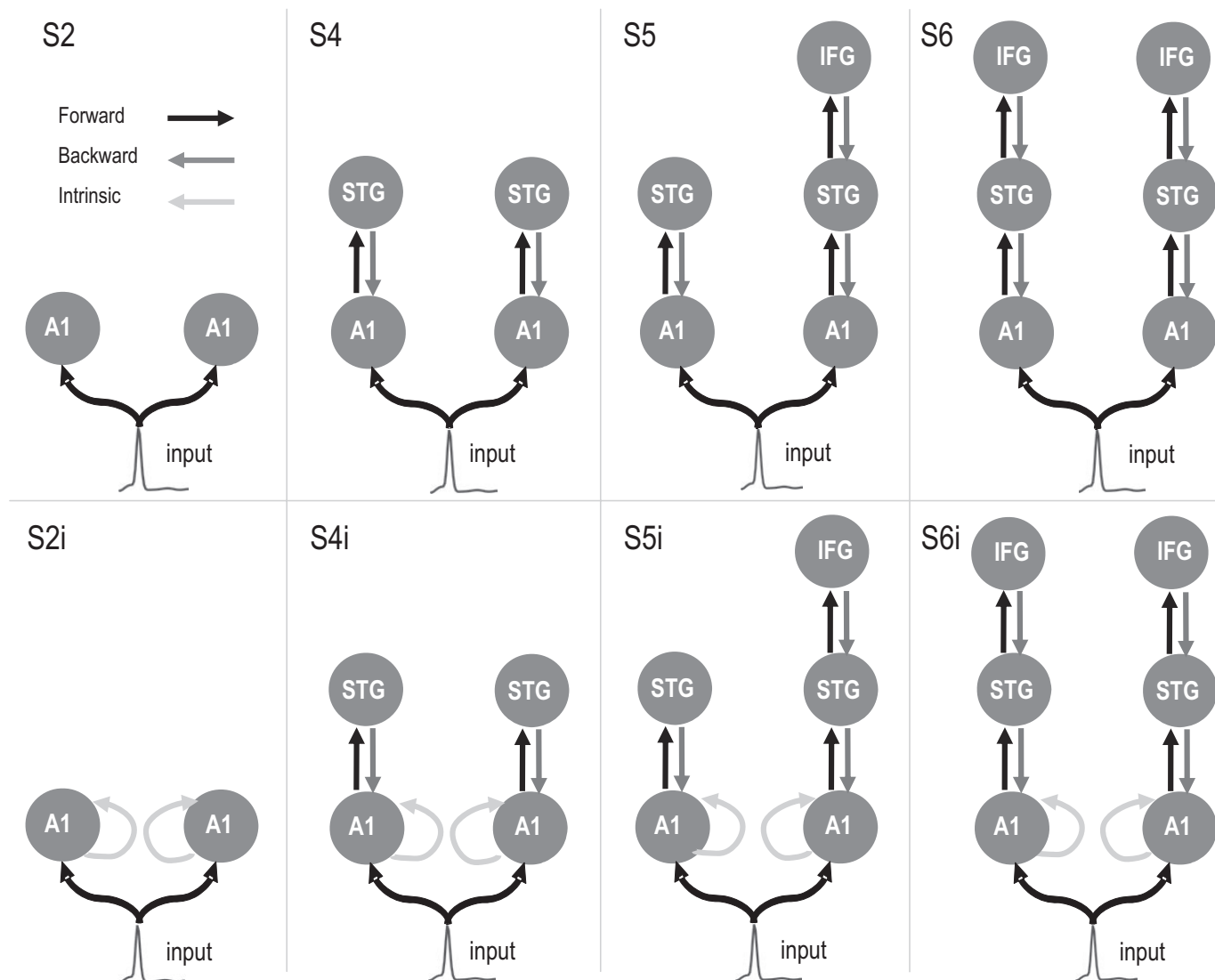


FIG. 2. Model specification. The sources comprising the networks are connected with forward (dark gray), backward (gray), or lateral (light gray) connections, all of which can show condition-specific changes. A1, primary auditory cortex; STG, superior temporal gyrus; IFG, inferior temporal gyrus. The 1st row of models, [S2, S4, S5, S6], allowed for changes in only extrinsic (forward and backward) connections, whereas the 2nd row [S2i, S4i, S5i, S6i] allowed for both extrinsic and intrinsic connectivity changes. Each column comprises similar models, which differ only in allowing for changes of intrinsic connectivity within A1 (i.e., adaptation).

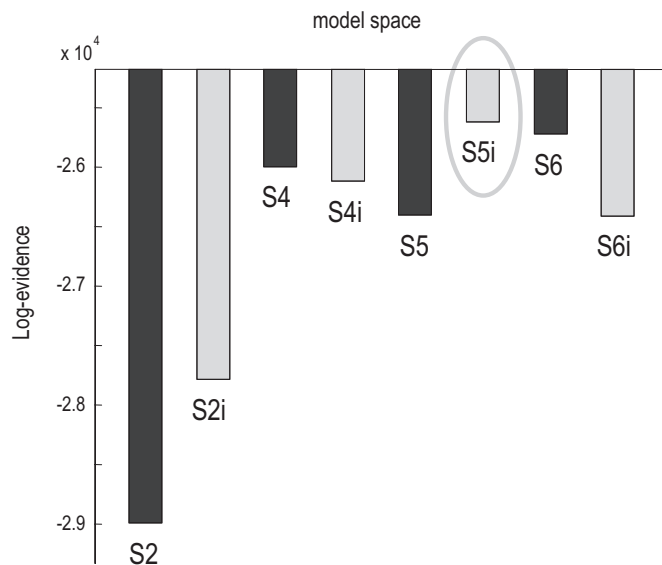


FIG. 3. Bayesian model selection. This graph shows the free-energy approximation to the log-evidence at the group level (i.e., pooled over subjects), for the 8 models. The best model is a 3-level hierarchical network comprising 5 interconnected cortical areas with local adaptation within primary auditory cortices (model: S5i).

connected cortical areas (emulating long-range connections between A1, STG, and the right IFG) and has condition-specific intrinsic connections at the level of the left and right A1 (emulating local adaptation). The fact that model S5i had the greatest log-evidence suggests that although local adaptation within the primary auditory cortices is important (compare S2 and S2i in Fig. 3), a much better explanation for the data rests on a combination of adaptation and plasticity in recurrent (i.e., forward and backward) connections among hierarchical cortical levels.

S6 was the second best model and S6i the third. The fact that there is less evidence for model S6i than for model S6 reflects an increase in model complexity (due to additional parameters encoding bilateral intrinsic connections within A1), which is not offset by the improvement in model fit. In short, our results show that model S5i has the best trade-off between model complexity and accuracy and is the most likely explanation of the data.

BETWEEN-SUBJECT CONSISTENCY. We tested for the consistency of the differences in model evidence across subjects using an ANOVA for repeated measures (with correction for nonsphericity) and the log-evidence as the dependent measure. The factors were the number of sources and changes in intrinsic connectivity. This analysis ensures that the differences in log-evidences are not driven by outliers and are large in relation to intersubject variability. We found an extremely significant main effect of hierarchical complexity, i.e., number of sources in the network [$F(2.23, 24.56) = 8.109, P < 0.001$]. A post hoc *t*-test showed the data were explained better by a hierarchical network model than a simple model with bilateral auditory sources.

RECONSTRUCTING SOURCE ACTIVITY. Figure 4 presents the results of model inversion quantitatively, for the best model (S5i), using the grand-mean response over subjects (using a prior variance of 4 mm² on dipole locations). Figure 4A shows

the prior locations of the dipolar sources overlaid in an MRI image of a standard brain. Figure 4B illustrates the reconstructed source activity generating the ERPs to standards and deviants for all three neuronal populations in each source. These differences are caused by, and only by, differences in coupling between and within sources. The average coupling gains and associated *P* values are shown against each connection. These represent average gains based on individual parameter estimates from subject-specific DCMs (the *P* values were computed using a one-sample *t*-test). These averages represent a scaling effect on comparing deviants with standards. For example, a coupling change of 0.52 from rSTG (right STG) to rIFG (right IFG) means that the effective connectivity decreased to 52% for deviants relative to standards and a coupling gain of 1.47 in the intrinsic connection within right A1 means a 47% increase in postsynaptic sensitivity during deviant trials. We found a consistent effect of coupling changes across group for intrinsic connections within left and right A1 (increase in connectivity; $P = 0.018$ and $P = 0.014$) and for the forward connection from STG to IFG (decrease in connectivity; $P = 0.015$). Backward connections from both STGs to their ipsilateral A1 show a trend for increases, which emulates a top-down effect on the lowest areas in the network.

The reconstructed source activities show transients that reflect both intrinsic and recurrent interactions among different neuronal populations. They show nicely how responses higher in the cortical hierarchy peak about 50 ms after responses in subordinate sources and, through backward connections, contribute to late components in lower sources. It is clearly difficult to interpret the detailed form of these evoked responses; however, because DCM is a generative or forward model we can isolate the contribution of specific connections by switching them on and off and looking at the ensuing responses. In the final analysis, we looked at the contribution of changes in intrinsic and extrinsic coupling to evoked responses using this approach.

DISSOCIATING RESPONSES TO DEVIANTS. Using the above-cited parameters of the DCM of grand-mean data, we integrated the predicted response (using the estimated latency and dispersion of subcortical input), while switching off extrinsic and intrinsic changes in connectivity. This enabled us to reconstruct observed responses (in the Cz electrode shown in Fig. 1) that would have been seen with just intrinsic adaptation or just extrinsic connection changes, respectively. We hoped to see that intrinsic differences were expressed early in the responses to deviant (cf. augmentation of the N1 component), whereas the contribution of extrinsic changes would be manifest later: during the MMN proper, when top-down predictions are available to make comparison between sensed and predicted bottom-up auditory input.

This is exactly what we observed. Figure 5A shows the simulated responses to standards and oddballs with intrinsic connection changes only. The corresponding difference waveform (Fig. 5D) emerges at about 100 ms and peaks at 200 ms. Conversely, when we allowed only extrinsic changes (Fig. 5B), the difference was expressed about 50 ms later, again peaking at about 150 ms (Fig. 5D). When both intrinsic and extrinsic changes operate together (Fig. 5C), the predicted (solid) and observed (dotted) responses are very similar (Fig. 5C).

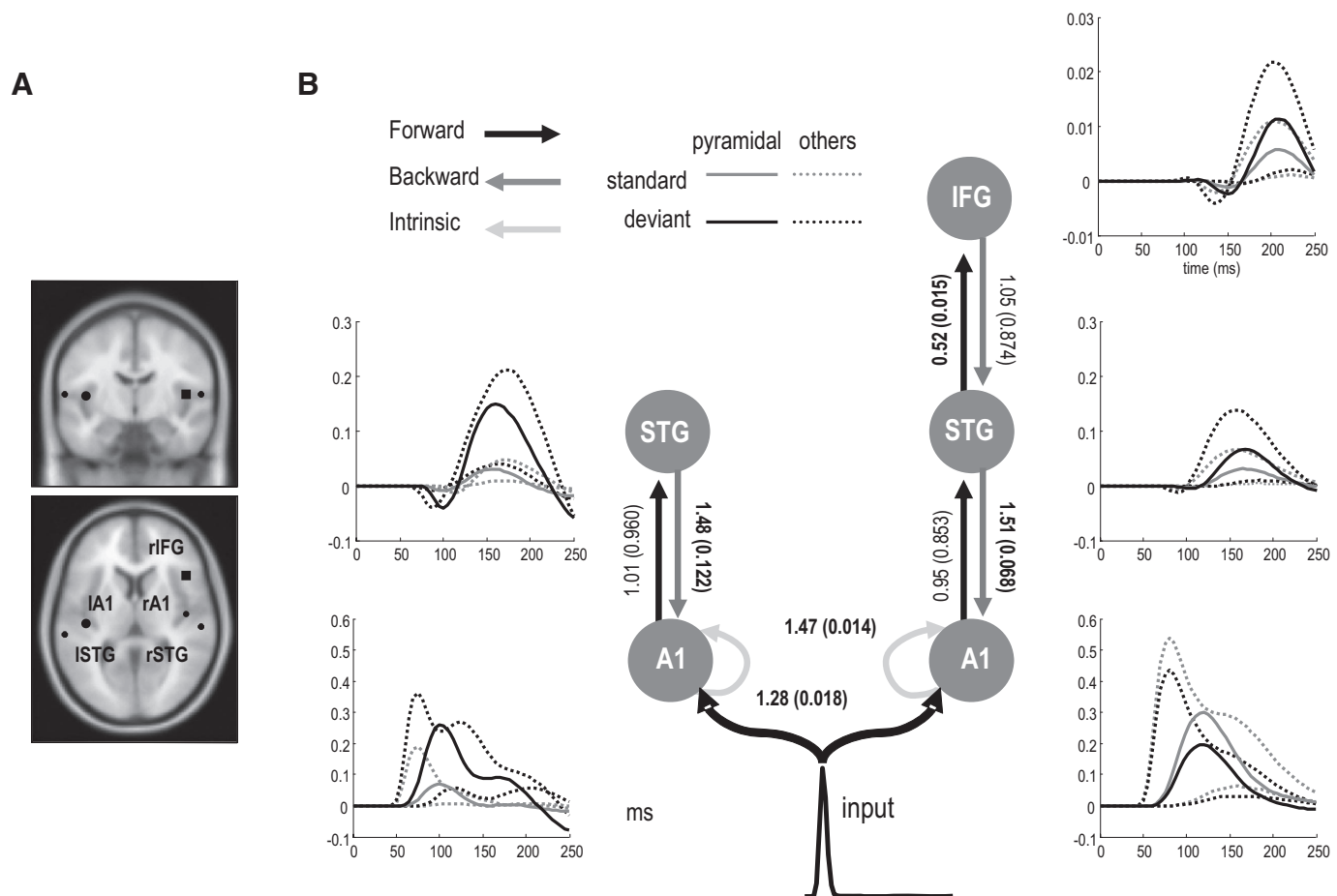


FIG. 4. Estimates of neuronal states and coupling changes for the grand-mean data under the best model (S5i). *A*: prior mean locations for the nodes or sources of activity (modeled as equivalent current dipoles), superimposed on a magnetic resonance image of a standard brain in Montreal Neurological Institute–defined space: lA1 [−42, −22, 7], rA1 [46, −14, 8], ISTG [−61, −32, 8], rSTG [59, −25, 8], lIFG [−46, 20, 8], and rIFG [46, 20, 8] in mm. *B*: reconstructed responses to deviants and standards at each source (the solid lines represent the depolarization of the pyramidal cell population that contributes to the observed ERP). The deviant response is expressed in all 3 populations in every source. There are widespread learning-related changes in most connections, expressed as coupling gains for deviants relative to standards. Mean coupling changes (i.e., averaged across subjects) and *P* values lie beside the connections in the graph.

DISCUSSION

In this study, we used DCM to explain ERPs to standards and deviants and tested different generative models of the observed responses. These models were motivated by theoretical accounts of the MMN that posit changes in synaptic efficacy in both extrinsic and intrinsic connections (e.g., Baldeweg 2006; Friston 2005). The distinction between effects mediated by plasticity in intrinsic and extrinsic coupling relates to established accounts of the MMN, specifically adaptation (Jääskeläinen et al. 2004) and model adjustment (Winkler et al. 1996).

Our model comparison addressed the importance of hierarchical sources in explaining responses to frequency deviants. These models allowed for changes in extrinsic connections alone among A1, STG, and IFG (i.e., forward and backward connections among sources) and were compared against a null model with no hierarchical sources or extrinsic connections. In addition, we assessed the importance of adaptation or changes in coupling that were intrinsic to auditory cortex. Although we could have limited our exploration of model space to a two-level hierarchy (A1 and STG), we included an IFG source, given some consistent findings in the literature (Doeller et al. 2003; Giard et al. 1990; Molholm et al. 2005; Optiz et al. 2002;

Rinne et al. 2000, 2005; Tse and Penney 2008). Bayesian model comparison revealed that the best model is a five-source network with intrinsic and extrinsic plasticity. Post hoc classical inference confirmed significant coupling changes, across subjects, in intrinsic connections within bilateral primary auditory cortices and extrinsically from right secondary auditory cortex to right frontal cortex. These are important findings because they provide direct evidence that responses to deviants are generated by self-organized dynamics, within a cortical hierarchy, that are mediated by changes in both extrinsic and intrinsic connectivity.

Functional anatomy of deviant responses

There is a controversy in the literature about the mechanisms underlying the MMN (see Näätänen et al. 2005 for a discussion). The adaptation hypothesis (Jääskeläinen et al. 2004) pertains to neurophysiological mechanisms, which do not necessarily entail an explicit computation. On the other hand, model adjustment speaks to some form of comparison (Näätänen and Winkler 1999; Winkler et al. 1996). Although the latter allows for adaptation effects (which the authors refer to as refractoriness), the adaptation hypothesis precludes a prediction or

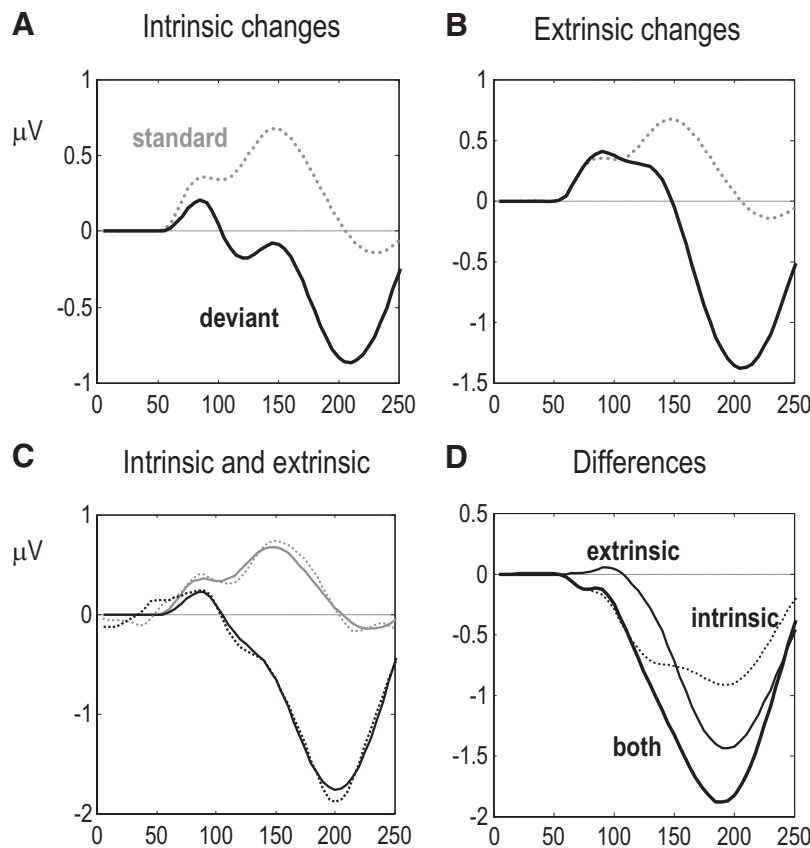


FIG. 5. Simulated ERPs at electrode C23 (Cz), predicted by the S5i DCM, using parameters optimized with the grand-mean average data. *A*: responses to standards (dotted line) and oddballs (solid lines) with changes in intrinsic connections but no extrinsic changes. *B*: responses with changes in extrinsic connections but no intrinsic changes. *C*: responses predicted under changes in both (solid lines) and those observed (dotted lines). *D*: difference waveforms for the 3 simulations in *A*, *B*, and *C*.

model-dependent contribution to the MMN. Predictive coding entails both adjustments to a generative model of stimuli and adaptation due to the increasing precision of predictions. We tested the relative contributions of model adjustment and adaptation by formulating them as network models with plastic changes in extrinsic (model-adjustment) and intrinsic (adaptation) connections. In doing this, we assumed that model adjustment rests on changes in top-down predictions to auditory cortex (acknowledging that model-adjustment could be mediated by mechanisms that are intrinsic to the auditory hierarchy). We showed that both model adjustment and adaptation contribute to the MMN and any associated N1 enhancement, consistent with predictive coding (Baldegeweg 2006; Friston 2005; Winkler 2007). The main conclusion from this study is that perceptual learning of stimulus trains changes both local and extrinsic coupling. This highlights the utility of physiologically constrained models of functional architectures to explain data and the usefulness of Bayesian model comparison.

Empirical evidence for frontal generators (Doeller et al. 2003; Giard et al. 1990; Molholm et al. 2005; Optiz et al. 2002; Rinne et al. 2000, 2005; Tse and Penney 2008) speaks to the potential role of distributed sources in generating the MMN. There are several interesting perspectives on the functional role of the frontal contribution to the processing of oddballs. An influential account suggests that the IFG mediates attentional switches or reorientation when salient environmental changes are detected (Näätänen 1990, 1992). The prefrontal cortex could be involved in the top-down modulation of a deviance-detection system, in the temporal cortices, as suggested by Escera et al. (2003). Although their study was based on modulation of the novelty-P3, Atenzia et al. (2002) presented

a discussion of the differences in MMN during sleep and wakefulness “that emphasizes the importance of the integration between bottom-up and top-down influences on sensory processing.” Alternative accounts of activity in the IFG suggest it might reflect contrast enhancement (Doeller et al. 2003; Optiz et al. 2002) or sensitivity tuning for change detection. It has also been suggested that the IFG is part of an inhibitory system, which allows sound changes to be ignored or suppressed, when these changes are irrelevant or small (Rinne et al. 2005; Tse and Penney 2008). It is interesting to note that our estimates of temporal to frontal connectivity decreased in deviant trials (Fig. 4). This is consistent with an inhibitory or suppressive role of prefrontal cortex when stimuli do not change. Having said this, it is very difficult to interpret changes in coupling parameters in functional terms (see following text).

Adaptation is a compelling hypothesis for the MMN; however, it does not explain empirical evidence suggesting that the MMN can be elicited by violation of abstract rules, such as complex interstimulus relationships (Näätänen et al. 2001; Paavilainen et al. 2001; Tervaniemi et al. 1994), irregularities in rhythms (Vuust et al. 2005), musical sequences (van Zuijlen et al. 2004), and violations in phoneme regularity (Näätänen 1997). Even a repeating sound in a randomly varying sequence of sounds will elicit MMN—termed “repetition negativity”—(Näätänen and Rinne 2002). Furthermore, an ascending tone pair in a sequence of descending tone pairs elicits an MMN (Saarinen et al. 1992), even though there is no stimulus repetition that could cause adaptation of a frequency-specific neuronal population. Given the tonotopic organization of auditory cortex, an MMN of this sort cannot be explained easily by local adaptation and seems to result from more complex mechanisms

involving more than one neuronal population and a fairly high level of perceptual synthesis.

DCM: inference and interpretation

It is tempting to interpret the changes in connectivity identified by DCM in quantitative terms; however, this is not as straightforward as it may appear. Both the real brain and the DCM can express complicated and nonlinear dynamics that do not always conform to simple intuitions—the effects of a change in an extrinsic connection may be expressed in a distributed and counterintuitive fashion. For example, in Fig. 4, the difference between standard and deviant responses in STG is mediated by both changes in bottom-up effects (from A1) and top-down effects from IFG. The bottom-up effects are partly due to significant increases in intrinsic connectivity within A1 and due partly to increases in backward connections from STG to A1 that modify recurrent dynamics. Furthermore, top-down effects are mediated by a reduction in the forward arc of reciprocal connections with IFG. The direction of change in any of these connections may cause changes in source activities that are not easy to predict; this is especially true of increases in backward connections that target both inhibitory and excitatory populations (see David et al. 2005).

Changes in intrinsic connectivity are simpler to interpret: The increase in intrinsic connectivity during deviants corresponds to an increase in the amplitude of postsynaptic responses to presynaptic inputs (Kiebel et al. 2007). Neuronal populations that are rendered less responsive through adaptation with repeated standard stimuli should therefore have a reduced sensitivity. Conversely, deviant stimuli will evoke greater postsynaptic responses in populations that have not adapted.

We used three different types of inference in this study. The first was Bayesian model comparison, which was used to identify the model that best explained our data (model S5i). This inference assumes that every subject's data were generated by the same model. In other words, if every subject had the same functional anatomy, model S5i had the most statistical evidence and this evidence was very strong. In some situations, it is possible the responses of different subjects are best explained by different models (cf. a random effect on models; KE Stephan, WD Penny, J Daunizeau, RJ Moran, and KJ Friston, unpublished observations). However, our question was not about intersubject variability. We simply pooled data over subjects to ask: Which was the best model, given a sample of normal subjects? To supplement this inference, we tested for the consistency of model evidence over subjects, using repeated-measures ANOVA and the log-evidence as a dependent measure. This revealed a main effect of hierarchy (number of sources). Finally, we selected the best model and tested for significant changes in connectivity, at the between-subject level, using the maximum a posteriori estimates from each subject. This allows for random effects on changes in connection strengths over subjects, under the best model. Here, we found consistent coupling changes within bilateral primary auditory cortex and from right secondary auditory cortex to right frontal cortex. In summary: 1) model S5i is the best, assuming each subject has the same neuronal architecture; and 2) allowing for random effects on coupling strengths, we found

significant changes in intrinsic (primary auditory cortex) and extrinsic (right auditory to prefrontal) connections.

The free energy (or log-evidence) is the optimum measure of model quality because it reflects not only the accuracy of DCM but also its complexity. Our results indicate that the dipoles added, beyond A1, improve model evidence. We did not assess the contribution of a second STG dipole (in relation to a unilateral STG). However, this is an interesting issue that could be addressed with model selection, by comparing models with and without a left STG. Indeed, Doeller et al. (2003) found that adding a left STG improves the variance explained by 6%. This is consistent with the finding that sequential activation in bilateral primary and secondary auditory cortex can explain late auditory-evoked potentials (Scherg and Von Cramon 1985). Other testable models include left-lateralized networks (left STG only or left IFG only models), which might be more likely in language MMN paradigms (cf. Näätänen et al. 1997). Finally, it would be interesting to investigate the role of adaptation (if any) in more refined paradigms that manipulate standards and deviants so that adaptation and model adjustment can be selectively suppressed (cf. Jacobsen and Schröger 2001; Maess et al. 2007; Rinne et al. 2006).

Modeling issues

DCM has been validated extensively in previous studies, both in terms of the source model (David and Friston 2003; David et al. 2004, 2005, 2006, 2008; Kiebel et al. 2006; Moran et al. 2007, 2008) and the specific network used here (Garrido et al. 2007a,b, 2008). However, a few points about DCM might be usefully reprised.

First, a crucial feature of DCM, and hypothesis-driven methods in general, is that one cannot test all possibilities. Therefore we can only make inferences about specific hypotheses, which embed our knowledge about functional anatomy. This study tested a number of models and one could argue that the winning model is a reasonable approximation to the real cortical network. However, any search of model space cannot be exhaustive; it selects only the best model among those considered. This means there might be other plausible network models. For example, it might be interesting to access the evidence of a more complex model with adaptation at the level of STG and/or IFG. Provided that there is a good motivation for extending the space of models, one can use Bayesian model comparison to select the best model in the new set.

DCM uses a conventional formulation of source localization (cf. equivalent current dipole [ECD] models; see Kiebel et al. 2006), but represents a departure from conventional inverse solutions to the EEG problem by using a full spatiotemporal forward model that embodies known constraints on the way EEG sources are generated (cf. Scherg and von Cramon 1985). Put simply, these constraints are that electrical activity in one part of the brain must be caused by activity in another (David et al. 2006). Conventional methods localize an active source associated with a specific peak at a given latency. In contrast, DCM models the activity as it evolves over all peristimulus times (≤ 250 ms in this study). Therefore the models considered here attempt to explain all the dynamics during that interval. DCM (and the brain) does not have any notion of response components or subcomponents such as N1, MMN, or P3a. DCM simply tries to explain the evoked responses that

unfold over a certain period of time, using a plausible neuronal model. Here, we were interested in explaining evoked responses (that subsume all components) in terms of changes in connectivity; for DCM these components are part of the same process. These include the MMN and any other components, such as the N1 or P3a. It is often difficult to disentangle the MMN and enhanced N1 components (see Jacobsen and Schröger 2001), especially when there is a large difference between standards and deviants. However, we were able to connect back to component-specific phenomenology in sensor space by using DCM to generate predicted responses with and without key mechanisms. This approach showed that the intrinsic (adaptation) mechanism produced an early (~ 120 ms) enhancement, as would be expected under the adaptation hypothesis and as suggested by attempts to dissociate sensory and comparison-based components of the MMN (Maess et al. 2007). Furthermore, a late negativity (~ 180 ms) was produced by extrinsic mechanisms that, under predictive coding, would be necessary for explicit change detection.

In summary, we used DCM to explain ERPs to standard and oddball stimuli and have framed our discussion in terms of the MMN. However, DCM has no notion of an MMN; it is trying to explain all the responses over peristimulus time and their differences. It would be disingenuous to claim we had identified the network responsible for the MMN. First, we have not identified the network (this would require an exhaustive search of model space); we have simply selected the best among a set of networks defining our hypothesis space. Second, the differences modeled are not restricted to the MMN (they may include N1 enhancement). In general, using DCM to model any specific waveform component calls on experimental design to isolate that component and make the subsequent inferences component specific.

Finally, it should be stressed that there is relatively little information in EEG measurements about the spatial location of sources (therefore changing the location priors does not change the DCM results very much). In contrast, there is an enormous amount of information about their orientation (see Kiebel et al. 2006). We therefore used tighter priors on the location than on the orientations (moments). We based our prior locations on the studies by Opitz et al. (2002) and Doeller et al. (2003) and chose bilateral STG and right IFG because they are reported consistently in the literature (see also Molholm et al. 2005). However, our model comparison would have been almost identical had we used different locations (we can infer this because, in the absence of priors, the conditional precision on the location parameters was in the order of centimeters not millimeters).

Conclusion

In conclusion, our results suggest that deviant responses cannot be attributed to local adaptation in the primary auditory cortices alone. This is consistent with a vast literature showing that temporal and frontal cortical sources are engaged by the MMN paradigm (Doeller et al. 2003; Jemel et al. 2002; Molholm et al. 2005; Opitz et al. 2002; Restuccia et al. 2005; Rinne et al. 2000, 2005). Our results support the idea that the MMN and related components rest on a more complex architecture involving interactions among hierarchical levels as well as local adaptation within the primary auditory cortices. As

noted by one of our reviewers, this is not really surprising if the DCMs model data containing both an N1 enhancement (adaptation) and MMN elicited by the frequency deviants. In other words, because DCM models both the N1 and MMN, any N1 enhancement is explained by the same mechanisms causing the MMN. In this sense, the mechanistic explanation afforded by DCM is not restricted to the MMN proper.

In fact, adaptation effects (neurons showing decreased activity due to stimulus repetition) as an explanation for the MMN had been discussed previously in terms of refractoriness (Näätänen 1992) or stimulus-specific adaptation (SSA) in single neurons in cat primary auditory cortex (Ulanovsky et al. 2003). Combined computational modeling and electromagnetic measurements by May et al. (1999) have also addressed local suppressive effects, or adaptation, and lateral inhibition—i.e., synaptic changes in horizontal connections intrinsic to an area. They have shown that the MMN can be explained by both neuronal adaptation and lateral inhibition. Although this work suggests that mechanisms of adaptation also underlie the MMN, it is distinct from the adaptation hypothesis, which claims that the generation of the observed response is due to adaptation alone and that there is no separate MMN (for a critical assessment see Näätänen et al. 2005).

Our results are consistent with the conjecture that the MMN reflects perceptual learning of standards, using predictive coding (Friston 2003, 2005; Garrido et al. 2007a)—i.e., adaptive changes in connectivity during perceptual discrimination of sounds (standards and deviants). These ideas are rooted in predictive coding models based on hierarchical Bayes (Rao and Ballard 1999). Predictive coding postulates that our perception of the world, under ambiguous sensory information, results from an interaction between our predictions, built from previous input, and the actual input from the environment. In this framework, ERPs reflect prediction error that is explained away (within-trial) by self-organizing neuronal dynamics during perception and is suppressed (between-trial) by changes in synaptic efficacy during perceptual learning. Differences in ERPs before and after learning (e.g., the MMN) can be interpreted as a failure to suppress prediction error, which can be explained quantitatively in terms of coupling changes among and within cortical regions. The repeated presentation of standards may render suppression of prediction error more efficient, leading to a reduction in evoked responses under repetition and the emergence of a deviant response, when a different stimulus is presented. The suppression of evoked responses, to a repeated event, is an ubiquitous phenomenon in neuroscience. It is seen at the level of single-unit responses (where it is referred to as repetition suppression; Desimone 1996) and is a long-standing observation in human neuroimaging [where it is often referred to as adaptation, e.g., cerebellar adaptation during motor repetitions (Friston et al. 1992) or repetition effects in visual studies (Henson et al. 2003)].

In models of predictive coding, units receiving sensory afferents are connected reciprocally to error units that effectively mediate a lateral inhibition (Friston 2008). A decrease in the excitability of A1 units may reflect a progressive increase in this inhibition with sensory learning and concomitant increases in the precision of prediction errors. This may be mediated by adaptation-like mechanisms. Changes in extrinsic connections may reflect changes in sensitivity to prediction error that is conveyed to higher levels. These higher levels

form predictions so that backward connections can provide contextual guidance to lower levels. In this view, the deviant responses represent a failure to predict bottom-up input and, consequently, a failure to suppress prediction error. The requisite change in architecture, during the implicit learning of standards, is expressed in terms of coupling changes among and within cortical regions, which we tried to quantify.

All the software necessary to implement these analyses is available as part of the SPM academic freeware (<http://www.fil.ion.ucl.ac.uk/spm>).

ACKNOWLEDGMENTS

We thank D. Bradbury for technical support, the volunteers for participating in this study, and the reviewers for help and thorough guidance in presenting this work.

GRANTS

This work was funded by Wellcome Trust grants to J. M. Kilner, S. J. Kiebel, and K. J. Friston and a Portuguese Foundation for Science and Technology grant to M. I. Garrido.

REFERENCES

- Atienza M, Cantero JL, Dominguez-Marín E. Mismatch negativity (MMN): an objective measure of sensory memory and long-lasting memories during sleep. *Int J Psychophysiol* 46: 215–225, 2002.
- Baldeweg T. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci* 10: 93–94, 2006.
- David O, Cosmelli D, Friston KJ. Evaluation of different measures of functional connectivity using a neural mass model. *NeuroImage* 21: 659–673, 2004.
- David O, Friston KJ. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* 20: 1743–1755, 2003.
- David O, Guillemain I, Saittel S, Rey S, Deransart C, Segebarth C, Depaulis A. Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol* 6: 2683–2697, 2008.
- David O, Harrison L, Friston KJ. Modelling event-related responses in the brain. *NeuroImage* 25: 756–770, 2005.
- David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ. Dynamic causal modelling of evoked responses in EEG and MEG. *NeuroImage* 30: 1255–1272, 2006.
- Desimone R. Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci USA* 93: 13494–13499, 1996.
- Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schröger E. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *NeuroImage* 20: 1270–1282, 2003.
- Escera C, Yago E, Corral M-J, Corbera S, Nunez MI. Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *Eur J Neurosci* 18: 2408–2412, 2003.
- Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.
- Friston K. Learning and inference in the brain. *Neural Networks* 16: 1325–1352, 2003.
- Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360: 815–836, 2005.
- Friston K. Hierarchical models in the brain. *PLoS Comput Biol* 4: e1000211, 2008.
- Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free-energy and the Laplace approximation. *NeuroImage* 34: 220–234, 2006.
- Friston KJ. Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage* 16: 513–530, 2002.
- Friston KJ, Frith CD, Passingham RE, Liddle PF, Frackowiak RS. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proc Biol Sci* 248: 223–228, 1992.
- Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *NeuroImage* 19: 1273–1302, 2003.
- Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM. The functional anatomy of the MMN: a DCM study of the roving paradigm. *NeuroImage* 42: 936–944, 2008.
- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci USA* 104: 20961–20966, 2007b.
- Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ. Dynamic causal modelling of evoked potentials: a reproducibility study. *NeuroImage* 36: 571–580, 2007a.
- Giard MH, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 27: 627–640, 1990.
- Henson RN, Goshen-Gottstein Y, Ganel T, Otten LJ, Quayle A, Rugg MD. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cereb Cortex* 13: 793–805, 2003.
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci USA* 101: 6809–6814, 2004.
- Jacobsen T, Schröger E. Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* 38: 723–727, 2001.
- Jansen BH, Rit VG. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern* 73: 357–366, 1995.
- Jemel B, Achenbach C, Müller BW, Ropcke B, Oades RD. Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topogr* 15: 13–27, 2002.
- Kiebel SJ, David O, Friston KJ. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage* 30: 1273–1284, 2006.
- Kiebel SJ, Garrido MI, Friston KJ. Dynamic causal modelling of evoked responses: the role of intrinsic connections. *NeuroImage* 36: 332–345, 2007.
- Maess B, Jacobsen T, Schröger E, Friederici AD. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *NeuroImage* 37: 561–571, 2007.
- May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Näätänen R. Frequency change detection in human auditory cortex. *J Comput Neurosci* 6: 99–120, 1999.
- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb Cortex* 15: 545–551, 2005.
- Näätänen R. *Attention and Brain Function*. Hillsdale, NJ: Erlbaum, 1992.
- Näätänen R, Alho K. Mismatch negativity—a unique measure of sensory processing in audition. *Int J Neurosci* 80: 317–337, 1995.
- Näätänen R, Jacobsen T, Winkler I. Memory-based or afferent process in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42: 25–32, 2005.
- Näätänen R, Rinne T. Electric brain response to sound repetition in humans: an index of long-term-memory-trace formation? *Neurosci Lett* 318: 49–51, 2002.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. “Primitive intelligence” in the auditory cortex. *Trends Neurosci* 24: 283–288, 2001.
- Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 125: 826–859, 1999.
- Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schröger E. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage* 15: 167–174, 2002.
- Paavilainen P, Simola J, Jaramillo M, Näätänen R, Winkler I. Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology* 38: 359–365, 2001.
- Penny WD, Stephan KE, Mechelli A, Friston KJ. Comparing dynamic causal models. *NeuroImage* 22: 1157–1172, 2004.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund H-J, Zilles K. Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13: 669–683, 2001.
- Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2: 79–87, 1999.
- Restuccia D, Della Marca G, Marra C, Rubino M, Valeriani M. Attentional load of the primary task influences the frontal but not the temporal generators of the mismatch negativity. *Cogn Brain Res* 25: 891–899, 2005.

- Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Näätänen R.** Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroImage* 12: 14–19, 2000.
- Rinne T, Degerman A, Alho K.** Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *NeuroImage* 26: 66–72, 2005.
- Rinne T, Särkkä A, Degerman A, Schroger E, Alho E.** Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Res* 1077: 135–143, 2006.
- Rosburg T, Haueisen J, Kreitschmann-Andermahr I.** The dipole location shift within the auditory evoked neuromagnetic field components N100m and mismatch negativity (MMNm). *Clin Neurophysiol* 115: 906–913, 2004.
- Sams M, Paavilainen P, Alho K, Näätänen R.** Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* 62: 437–448, 1985.
- Scherg M, Von Cramon D.** Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol* 62: 32–44, 1985.
- Sussman E, Winkler I.** Dynamic sensory updating in the auditory system. *Brain Res Cogn Brain Res* 12: 431–439, 2001.
- Tervaniemi M, Maury S, Näätänen R.** Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 5: 844–846, 1994.
- Tiitinen H, May P, Reinikainen K, Näätänen R.** Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372: 90–92, 1994.
- Tse CY, Penney TB.** On the functional role of temporal and frontal cortex activation in passive detection of auditory deviance. *NeuroImage* 41: 1462–1470, 2008.
- Ulanovsky N, Las L, Nelken I.** Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6: 391–398, 2003.
- van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M.** Grouping of sequential sounds—an event-related potential study comparing musicians and non-musicians. *J Cogn Neurosci* 16: 331–338, 2004.
- Vuust P, Pallesen KJ, Bailey C, van Zuijen TL, Gjedde A, Roepstorff A, Østergaard L.** To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage* 24: 560–564, 2005.
- Wager TD, Keller MC, Lacey SC, Jonides J.** Increased sensitivity in neuroimaging analysis using robust regression. *NeuroImage* 26: 99–113, 2005.
- Winkler I.** Interpreting the mismatch negativity (MMN). *J Psychophysiol* 21: 147–163, 2007.
- Winkler I, Karmos G, Näätänen R.** Adaptive modelling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res* 742: 239–252, 1996.