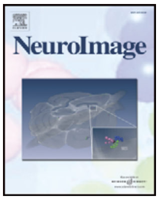
NeuroImage 42 (2008) 936–944

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



The functional anatomy of the MMN: A DCM study of the roving paradigm

Marta I. Garrido [a,c,*], Karl J. Friston [a], Stefan J. Kiebel [a], Klaas E. Stephan [a], Torsten Baldeweg [b], James M. Kilner [a]

[a] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, UK
[b] Developmental Cognitive Neuroscience, Institute of Child Health, University College London, UK
[c] Department of Psychology, University of California, Los Angeles, USA

ARTICLE INFO

Article history:
Received 4 March 2008
Revised 1 May 2008
Accepted 11 May 2008
Available online 20 May 2008

ABSTRACT

Using dynamic causal modelling (DCM), we have presented provisional evidence to suggest: (i) the mismatch negativity (MMN) is generated by self-organised interactions within a hierarchy of cortical sources [Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36, 571–580] and (ii) the MMN rests on plastic change in both extrinsic (between-source) and intrinsic (within source) connections (Garrido et al., under review). In this work we re-visit these two key issues in the context of the roving paradigm. Critically, this paradigm allows us to discount any differential response to differences in the stimuli per se, because the standards and oddballs are physically identical. We were able to confirm both the hierarchical nature of the MMN generation and the conjoint role of changes in extrinsic and intrinsic connections. These findings are consistent with a predictive coding account of repetition–suppression and the MMN, which gracefully accommodates two important mechanistic perspectives; the model-adjustment hypothesis [Winkler, I., Karmos, G., Näätänen, R., 1996. Adaptive modelling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res. 742, 239–252; Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125, 826–859; Sussman, E., Winkler, I., 2001. Dynamic sensory updating in the auditory system. Brain Res. Cogn Brain Res. 12, 431–439] and the adaptation hypothesis [May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G., Näätänen, R., 1999. Frequency change detection in human auditory cortex. J. Comput. Neurosci. 6, 99–120; Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. U. S. A. 101, 6809–6814].

Introduction

Novel events, or oddballs, embedded in a stream of repeated events, or standards, produce a distinct response that can be recorded non-invasively with electrophysiological techniques such as electroencephalography (EEG). The mismatch negativity (MMN) is one of the ERP components elicited by any discriminable violation in the acoustic regularity. The MMN is believed to be an index of automatic change detection governed by a pre-attentive sensory memory mechanism (Näätänen, 1990). Despite being the subject of much research, the mechanisms behind MMN generation remain a topic of much debate. Recently, we provided evidence that the mechanisms underlying the MMN can be considered within a hierarchical inference or predictive coding framework (Garrido et al., 2007). Within this account the MMN is interpreted as a failure to suppress prediction error, which can be explained quantitatively in terms of coupling changes among and within cortical regions. The predictive coding framework encompasses two previous hypotheses that have been debated in the literature; the adaptation hypothesis (May et al., 1999; Jääskeläinen et al., 2004) and the model-adjustment hypothesis (Winkler et al., 1996; Näätänen and Winkler 1999; Sussman and Winkler, 2001). While 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 adjustments to a generative model of stimulus trains 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

[*] Corresponding author. UCLA Department of Psychology, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA.
E-mail address: migarrido@ucla.edu (M.I. Garrido).

1053-8119/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.05.018

(model-adjustment) and intrinsic (adaptation) connections. We show that both model-learning and adaptation contribute to the MMN, consistent with predictive coding or model based explanations (Friston et al., 2006; Winkler, 2007).

It has been suggested that stimulus repetition engenders an echoic memory trace, which compares preceding and current stimuli (Näätänen, 1992). The MMN increases with the number of repetitions of a standard stimulus and is believed to reflect the strength of this trace (Sams et al., 1993). Repetition in roving paradigms, which are characterised by sporadic changes in the frequency of a repeating tone, enhance a slow positive wave from 50 to 250 ms post-stimulus in the standard ERP; the repetition positivity (RP) (Baldeweg et al., 2004). Both RP and MMN increase with repetitions of standards, suggesting that these are the ERP correlates of sensory memory formation and are (the same) electrophysiological signatures of sensory learning (Haenschel et al., 2005).

In the predictive coding framework (see also Friston 2005; Baldeweg 2006), evoked responses, corresponding to prediction error, drive perceptual inference (within-trial) and changes in connectivity (between trials) so that prediction error is suppressed with learning. Previously, we used dynamic causal models (DCMs) to explore network models underlying mismatch or oddball responses (Garrido et al., under review). This was achieved by explaining differences in the ERP evoked by standard and deviant tones on the basis of plastic changes within a cortical network. DCM models every time-bin and every channel in a single analysis and attempts to explain differences in the evoked responses, including the MMN and other components such as the N1 (Näätänen et al., 2005), in terms of changes in connection strengths. Our previous study employed a classical oddball paradigm, which meant that any differences in between ERPs evoked by standard and deviant tones could have been driven by a stimulus-specific N1 response as well a pure MMN response. Here we used a roving paradigm, where there were no acoustic differences between the standard and deviant. This ensured that any differences could not be explained stimulus-specific differences in the N1 contribution and enabled us to model the MMN per se.

In short, the key contributions of this study are firstly, to demonstrate the validity of the mechanism for MMN generation that we have proposed previously (Garrido et al., under review), and secondly, to show that the ensuing mismatch responses are due to learning and not to stimuli differences *per se*.

Methods

Stimuli

We studied twelve healthy volunteers aged 24–34 (4 female). Each subject gave signed informed consent before the study, which proceeded under local ethical committee guidelines. Subjects sat in front of a desk in a dimly illuminated room. Electroencephalographic activity was measured during an auditory roving 'oddball' paradigm (see Fig. 1a). The stimuli comprised a structured sequence of pure sinusoidal tones, with a roving, or sporadically changing tone. This paradigm resulted from few modifications to that used in Haenschel et al. (2005), originally design by Cowan et al. (1993). Within each stimulus train, all tones were of one frequency and were followed by

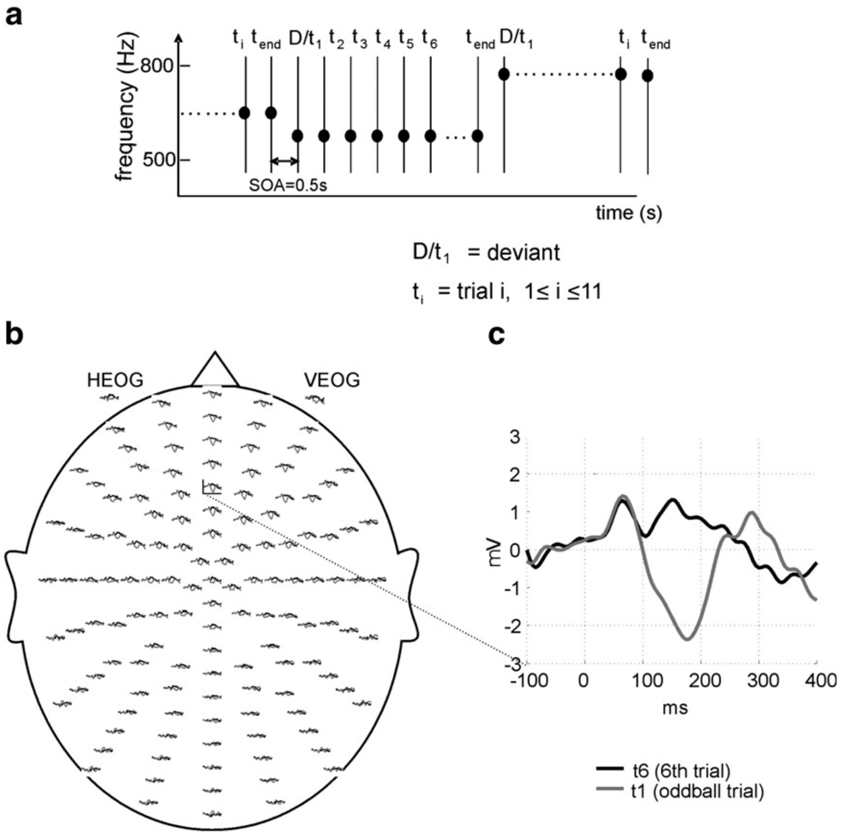


Fig. 1. Design and responses elicited in a roving paradigm. (a) Stimulus design is characterised by a sporadically changing standard stimulus. The first presentation of a novel tone is a *deviant* (D/t$_1$) that becomes a *standard*, through repetition (t$_6$). However, in this paradigm, deviants and standard have exactly the same physical properties. (b) Grand mean (averaged over all subjects) ERP responses to the sixth tone presentation, the established "standard" (t$_6$ in black) and deviant tone (D/t$_1$, in grey) overlaid on a scalp-map of 128 EEG electrodes. (c) enlarged ERP responses to the standard and deviant tones at channel C21 (fronto-central) where the MMN response peaks at about 180 ms from change onset.

a train of a different frequency. The first tone of a train was a deviant, which eventually became a standard after few repetitions. So deviants and standards have exactly the same physical properties, differing only in the number of times they have been presented. The number of times the same tone was presented varied pseudo-randomly between one and eleven. The probability that the same tone was presented once or twice was 2.5%; for three and four times the probability was 3.75% and for five to eleven times it was 12.5%. The frequency of the tones varied from 500 to 800 Hz in random steps with integer multiples of 50 Hz. Stimuli were presented binaurally via headphones for 15 min. The duration of each tone was 70 ms, with 5 ms rise and fall times, and the inter-stimulus interval was 500 ms. About 250 deviant trials (first tone presentation) were presented to each subject. About 250 deviant trials (first tone) and about 200 standards (sixth tone) were presented to each subject. Each subject adjusted the loudness of the tones to a comfortable level, which was maintained throughout the experiment. The subjects performed a distracting visual task and were instructed to ignore the sounds. The task consisted of button-pressing whenever a fixation cross changed its luminance, which occurred pseudo-randomly every 2 to 5 s (and did not coincide with auditory changes).

Data acquisition and pre-processing

EEG was recorded with a *Biosemi* system with 128 scalp electrodes. Data were recorded at a sampling rate of 512 Hz. Vertical and horizontal eye movements were monitored using EOG (electro-oculograms) electrodes. Pre-processing and data analysis were performed with SPM5 (http://www.fil.ion.ucl.ac.uk/spm/). The data were epoched offline, with a peri-stimulus window of −100 to 400 ms, down-sampled to 200 Hz, band-pass filtered between 0.5 and 40 Hz and re-referenced to the nose. The method used for artefact removal was robust averaging. Robust averaging is a standard averaging routine. It is an iterative algorithm that produces the best estimate of the average by weighting data points as a function of their distance from an estimate of the mean for each iteration (*c.f.* Wager et al., 2005). Trials were sorted in terms of tone repetition. In other words, trials one to eleven correspond to the responses elicited after one to eleven presentations of the same tone, collapsed across the whole range of frequencies. Trial one is the oddball, or the deviant trial. Two subjects were excluded from the analysis due to artefacts and another due to an undetectable MMN. Data were transformed into scalp-map images (see Fig. 2a). These were obtained after linear interpolation and smoothing (at FWHM 6:6:4 a.u.) of the difference wave response between the first presentation and the sixth presentation. For computational expediency, DCMs (see below) were computed on a reduced form of data that corresponded to eight channel mixtures or spatial modes. These were the eight principal modes of a singular value decomposition (SVD) of the channel data between 0 and 250 ms, over trial types of interest. The use of eight principal eigenvariates explained on average 80% of the variance in the data across the group (and more than 70% of the data in every subject).

Dynamic causal modelling

Dynamic causal model (DCM) was originally developed for connectivity analysis of fMRI (Friston et al., 2003) and M/EEG data (David et al., 2006). Most approaches to connectivity analysis of M/EEG data use functional connectivity measures such as coherence or temporal correlations, which establish statistical dependencies between two time-series. However, there are certain cases where causal interactions are the focus of interest. In these situations, DCM is particularly useful, because it estimates effective connectivity (the influence one neuronal system has over another), under a perturbation, or stimulus. DCM provides an account of the interactions among cortical regions and allows one to make inferences about the parameters of the system and investigate how these parameters are influenced by experimental factors. DCM furnishes spatiotemporal, generative or forward models for evoked responses as measured with EEG/MEG (David et al., 2006; Kiebel et al., 2006), and provides an important advance over conventional analyses of evoked responses because it places natural constraints on the inversion; namely, activity in one source has to be caused by activity in another. DCMs for MEG/EEG use neural mass models (David and Friston, 2003) to explain source activity in terms of the ensemble dynamics of interacting inhibitory and excitatory subpopulations of neurons, based on the model of Jansen and Rit (1995). The active sources are interconnected according to the connectivity rules described in (Felleman and Van Essen, 1991) and conform to a hierarchical model of intrinsic and extrinsic connections within and among multiple sources as described in David et al. (2005) and Kiebel et al. (2007).

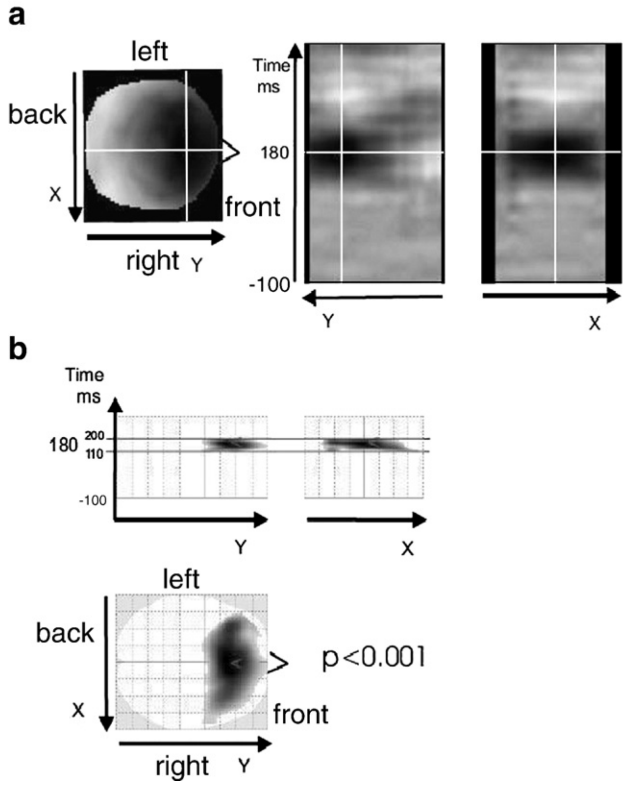


Fig. 2. 3D spatiotemporal characterisation of the grand mean difference wave: comparison between the first and the sixth presentations; the "deviant" and the "standard", respectively. This analysis searched for differences over 2D sensor-space (interpolated from 128 channels) and all peri-stimulus times (−100 to 400 ms). (a) The differential response shows a negative peak at about 180 ms over the frontal and central areas. (b) SPM showing where, over subjects, there is a significant negative difference at the between-subject level (p<0.001 uncorrected). Significant effects were found over temporal and frontal areas in the range 110 to 200 ms peaking at 180 ms (see marker).

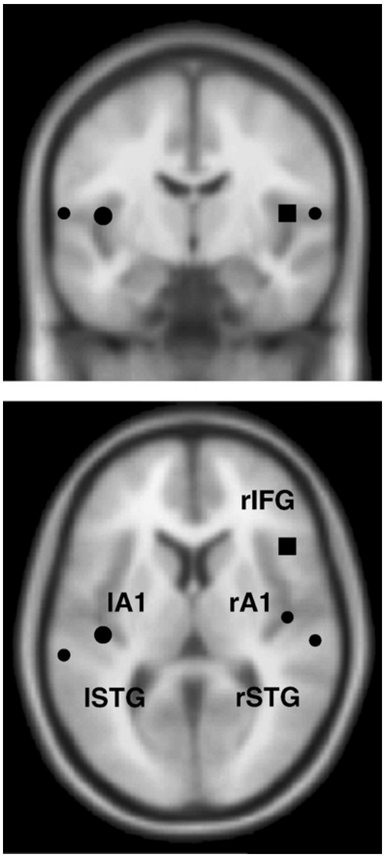


Fig. 3. Prior locations for the nodes in the models. Sources of activity were modelled as equivalent dipoles. Their prior mean locations: lA1 [−42, −22, 7], rA1 [46, −14, 8], lSTG [−61, -32, 8], rSTG [59, −25, 8], lIFG [−46, 20, 8], rIFG [46, 20, 8] in mm are superimposed in an MRI of a standard brain in Montreal Neurological Institute (MNI) space.

By taking the marginal likelihood over the conditional density of the model parameters, one can estimate the probability of the data, given a particular model. This is known as the marginal likelihood or evidence and can be used to compare and select the best model amongst alternative models. We have previously used DCM to explain ERPs to standards and deviants using a classical paradigm (Garrido et al., under review). Differences in the ERP to standards and deviants were modelled in terms of changes in synaptic connections within and between hierarchically organised cortical sources. As in this study, our model space was motivated by previous accounts of the MMN, specifically *adaptation* (Jääskeläinen et al., 2004), and *model-adjustment* (Winkler et al., 1996). Model comparison addressed hierarchical implementations of multiple-level network models ranging from one to three levels. These models allowed for changes in extrinsic connections alone (i.e., forward and backward connections among A1, STG and IFG) or in combination with changes in intrinsic connections at the level of A1. Bayesian model comparison showed that the best model was a five-source network with both intrinsic and extrinsic plasticity. Here, we investigate whether the same model could explain the MMN elicited in a roving paradigm, where differential N1 components can be discounted.

Model specification

DCM is a hypothesis-driven method: it does not explore all possible models but tests specific mechanistic hypotheses, defined in terms of specific connectivity models. Bayesian model selection of DCMs can provide evidence in favour of one model relative to others. The results of a DCM analysis depend explicitly upon the models evaluated. Our network architectures were motivated by the results of previous studies of MMN generators (Rinne et al., 2000; Opitz et al., 2002; Doeller et al., 2003; Grau et al., 2007; Garrido et al., 2007; under review). These studies suggest bilateral sources located in the superior temporal gyrus (STG), and inferior frontal gyrus (IFG), which are usually stronger, and found more consistently in the right hemisphere.

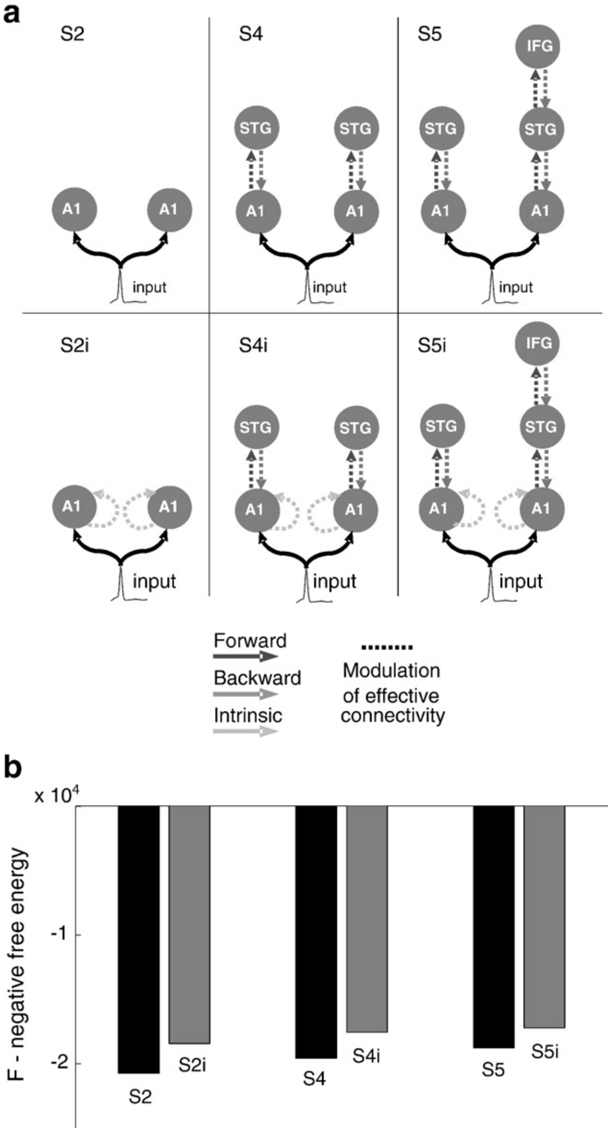


Fig. 4. Model specification and Bayesian model comparison for the six networks tested. (a) The sources comprising the networks: A1: primary auditory cortex, STG: superior temporal gyrus and IFG: inferior temporal gyrus are connected with forward (dark grey), backward (grey) and intrinsic (light grey) connections. The first row of models, [S2, S4, S5], allowed for learning-related changes in only extrinsic (forward and backward), while the second row [S2i, S4i, S5i] allowed for conjoint changes in extrinsic and intrinsic connections. Each column is filled with two similar network models, which differ only in allowing for modulations of intrinsic connectivity within A1. From one column to the next we have increased the number of active sources and connected them reciprocally (with forward and backward connections). (b) The graph shows the free-energy approximation to the log-evidence at the group level, *i.e.*, pooled over subjects, for the six models. The best model is a 3-level hierarchical network, comprising five interconnected cortical areas; allowing for local adaptation within primary auditory cortices and plastic changes in extrinsic connections (model- S5i).

Our models attempt to explain the generation of each individual response (i.e., responses to each tone presentation). Therefore, left and right A1 were chosen as cortical input stations for processing the auditory information. Doeller et al. (2003) identified sources for the differential response, with fMRI and EEG measures, in both left and right STG and IFG. Here, we have modelled each active source, *i.e.*, each node in the network, with a single equivalent current dipole (ECD). We 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 mapping cortical sources onto measured signals was parameterised in terms of the location and orientation of each dipole source (see Kiebel et al., 2006 for details). This employs the electromagnetic forward model solutions encoded in the fieldtrip software (http://www2.ru.nl/fcdonders/fieldtrip). The coordinates reported by Opitz et al. (2002) (for STG and IFG) and Rademacher et al. (2001) (for left and right A1) 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 MNI space using the algorithm described in (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach) (see Fig. 3). The moment parameters had 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. In the models considered, all extrinsic connections were reciprocal and the exogenous (subcortical auditory) input entered bilaterally to primary auditory cortices (A1).

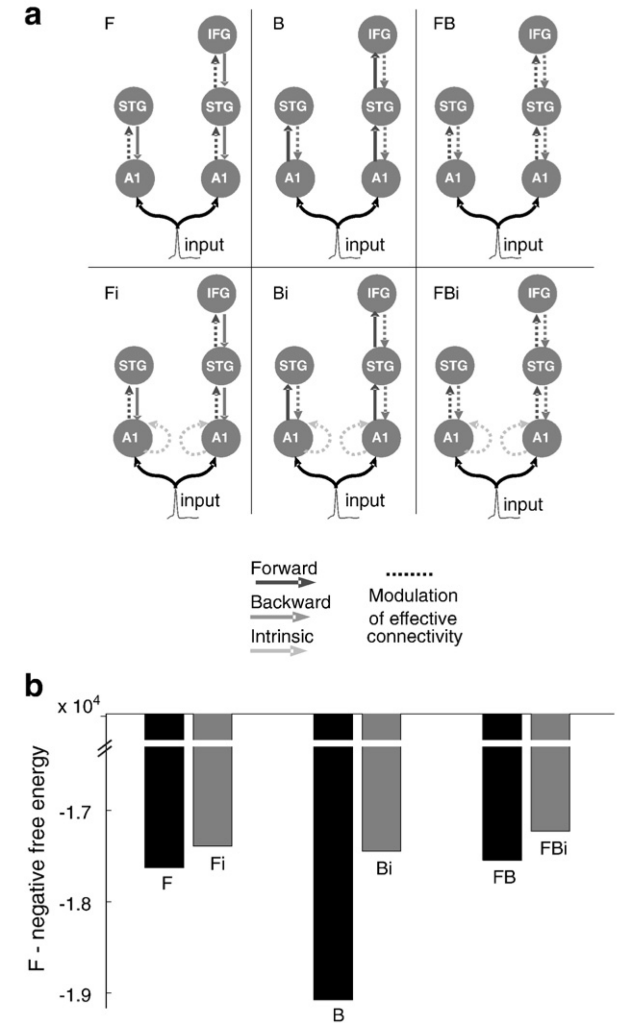


Fig. 5. Model specification and Bayesian model comparison for the six variants of model S5i. (a) Six models comprising three hierarchical cortical levels. Bilateral A1 are reciprocally connected with bilateral STG, and right STG is reciprocally connected with right IFG. The first row of models, [F, B, FB], allowed for learning-related changes in only extrinsic connections: forward, backward and conjoint forward and backward connections, respectively. The second row [Fi, Bi, FBi] allowed for conjoint extrinsic and intrinsic (within A1) connections. Each column shows similar network models, which differ only in allowing for changes of intrinsic connectivity within A1. (b) The graph shows the free-energy approximation to the log-evidence at the group level, *i.e.*, pooled over subjects, for the six models. The best model is FBi which allows for modulations of all extrinsic and intrinsic connections. This is in fact exactly the same as the parent model S5i in which all connections could change. Models Fi and FBi are better than Bi.

DCMs

Six models were specified by their architectures (see Fig. 4a). These models cover different mechanisms for the MMN generation, including the: *the adaptation hypothesis* (Jääskeläinen et al. 2004),[1] *model-adjustment* (Winkler et al., 1996) and combinations of the two (for details on model specification see Garrido et al., under review; and for a critical assessment see Näätänen et al., 2005). The model search started with the most parsimonious model, S2, (a one-level hierarchical model comprising two nodes in the left and right primary auditory cortex, A1), and increased in their complexity, in terms of hierarchical levels, number of sources and changes in intrinsic connectivity. The inclusion of nodes and connections to the initial model culminated in a non-symmetric three-level hierarchical model that included bilateral A1 and STG, and right IFG. All models can therefore be considered as a sub-model of the last, S5i). Our simplest model, S2, is a two source network corresponding to the hypothesis that the ERPs to standards and deviants are generated by bilateral activity in A1. This model is naïve in the sense that it does not support changes in connectivity or consequent changes in ERPs. Model S2i is similar to S2 but allows for coupling changes within A1. Here, we hypothesise that differences between responses to standards and deviants are caused by changes in A1 activity due to modulations of intrinsic connections within this area. This model resembles *the adaptation hypothesis*. Here, we attempted to model adaptation effects with changes in intrinsic connectivity (within A1) — model S2i. This allows for stimulus-specific adaptation (SSA — Ulanovsky et al., 2003), where repetitive auditory events adapt feature-specific neurons cumulatively. S2 does not have the latitude to model this because its intrinsic interactions are fixed. This makes it a naïve model because it cannot explain any ERP differences that are present in the data. Model S4 is a second-level hierarchical model comprising four sources. It builds on S2 through addition of left and right superior temporal gyrus (STG) source (connected reciprocally through forward and backward connections to ipsilateral A1). This was motivated by the general principle of *reciprocity* in cortico-cortical connections: two areas are linked through anti-

[1] The adaptation hypothesis postulates that the MMN arises predominantly from post-synaptic mechanisms, i.e. spike-frequency adaptation due to increase in calcium-dependent potassium conductances, leading to slow afterhyperpolarizing currents (c.f. May et al., 1999). Here, we model similar adaptive effects through changes in post-synaptic density parameters (see Kiebel et al., 2007 for details).

parallel or bidirectional pathways. See for example Rockland and Pandya (1979) or Felleman and Van Essen (1991). Model S4i is analogous to S4 with additional intrinsic connections within A1. A third-level hierarchical model S5, comprising five sources, included a right inferior frontal gyrus (IFG) source. Right STG was reciprocally connected with ipsilateral IFG. Model S5i is like S5 but has additional self-connections within A1. Models S4 and S5 embody mechanisms that are consistent with *the model-adjustment hypothesis*. Models S4i and S5i cover the hypothesis that both local adaptation, within primary auditory cortex, and interactions within a temporofrontal network underlie the generation of the MMN.

Statistical analysis

Statistical analyses in this paper were based on model comparison. Model m is inverted by optimising a variational free-energy bound, F, on the model-evidence to provide the conditional density of the model parameters, $p(\theta|y,m)$, and the models' evidence, $p(y|m)$, for model comparison. Specifically, inversion of a DCM corresponds to approximating the posterior probability of the parameters using variational Bayes as described in Friston (2002). The aim is to minimise a free-energy bound on the log-evidence, with respect to a variational density, $q(\theta)$. When the free-energy is minimised; $q(\theta)=p(\theta|y,m)$ and the free-energy $F=-\ln p(y|m)$ approximates the negative marginal log-likelihood or negative log-evidence. After convergence the variational density is used as an approximation to the desired conditional density and the log-evidence is used for model comparison.

One often wants to compare different models and select the best before making statistical inferences on the basis of the conditional density. The best model, given the data, is the one with highest log-evidence, $\ln p(y|m)$ (assuming a uniform prior over models). Given two models m_1 and m_2 one can compare them by computing 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 greater than about three (*i.e.*, their relative likelihood is more than 20:1) then one asserts there is strong evidence in favour of the first model. This formalism is suitable for comparing different models of a given data set, for instance data acquired from a single subject. However, one may wish to select the model that best explains multiple data sets, *i.e.*, the best model at the group level. Assuming each data set is independent of the others (*i.e.*, all subjects are measured independently), we can simply multiply the marginal likelihoods or, equivalently, add the log-evidences from each subject to obtain the log-evidence for a given model over subjects (Garrido et al., 2007).

Results

Learning the acoustic environment through stimulus repetition changes connectivity within and between hierarchically organised cortical areas. This analysis comprised three parts: (i) confirmation that there is a significant differential response (MMN) between the first and sixth tone presentation; (ii) model selection to identify the most likely number and hierarchical deployment of sources causing these responses and (iii) hypotheses or model testing to establish that the MMN is mediated by changes in both extrinsic and intrinsic connectivity (under the best hierarchical model).

Mismatch responses due to repetition effects

An initial analysis was performed to confirm the presence of a MMN response in our roving paradigm. We recorded data from 128 EEG sensors while subjects listened to trains of pure tones. Each stimulus train was comprised of a sequence of equal tones, and was followed by another stimulus train of a different frequency. Within a stimulus train, each tone was presented between one and eleven times before changing. The first presentation of a tone with a different frequency from the preceding tone was defined as a deviant (see Methods and Fig. 1a for details on experimental design). Fig. 1b shows the grand mean responses (i.e., averaged across subjects) to first tone presentation; the deviant or oddball trial (D/t_1, in grey), and responses to the sixth presentation (t_6 in black), when we assume a “standard” response has been attained. This assumption is based on the ERP forms for D/t1 and t_6 shown in Fig. 1c. Indeed, a MMN response was found over the frontal and temporal electrodes, peaking at about 180 ms from change onset, which is consistent with previous studies (Cowan et al., 1993; Baldeweg et al., 2004). Fig. 1c shows the enlarged responses to the deviant and the sixth tone presentation, or “standard” at a fronto-central electrode (C21), where the MMN was more evident.

Fig. 2 shows a 3D spatiotemporal characterisation of the grand mean difference wave response, using statistical parametric mapping to compare the first and the sixth presentations; the “deviant” and the “standard”, respectively. This analysis searched for differences over 2D sensor-space and all peristimulus time [−100, 400]. The scalp topography at any time-bin was interpolated from 128 channels and smoothed. Fig. 2a shows the intensity of the differential response and that its negative peak occurs at about 180 ms over the frontal and central areas. Fig. 2b shows the corresponding statistical parametric map (SPM) where, over subjects, there is a significant negative difference across subjects ($p<0.001$ uncorrected). This SPM showed a significant MMN over frontal areas between 110 and 200 ms, with maximum at 180 ms.

Underlying connectivity models of the MMN

Next, we tested different hierarchical models that represent specific mechanistic hypotheses about MMN generation: *adaptation* (mapped onto S2i-model), *model-adjustment* (S4- and S5-models) and *predictive coding* (S4i- and S5i-models). Models S4i and S5i could also be regarded as an *adaptation* models, if the differences in the ERPs to standards and deviants were driven by modulations in the intrinsic connections only. Both responses, ERPs to standards and deviants, were explained by the same model in these analyses. The differences in the ERPs; *i.e.*, the MMN, are explained in terms of coupling changes within and among the cortical areas of the underlying network model. The aim of these analyses was to assess whether we could replicate our previous results using classical oddball paradigms (Garrido et al., 2007; under review). Indeed, the best model was the same for the two independent experiments, model S5i. The models illustrated in Fig. 4a differed in terms of their nodes and in the connections which could show putative learning-related changes, *i.e.*, differences between listening to standard or deviant tones. Models S4 and S5 allowed for changes in all extrinsic (forward and backward) connections, which map to hypotheses that differences in ERPs to standards and deviants are due to plasticity in extrinsic connections; and models S4i and S5i allowed for changes in the same extrinsic connections plus changes in intrinsic connections within left and right A1. These models map to hypotheses that differences in ERPs are due to conjoint coupling changes in extrinsic and intrinsic connections. An ANOVA test for repeated measures on the free-energy (an approximation to each model's log-evidence) revealed a main effect of *source number* ($p<0.04$) and a main effect of *intrinsic connectivity* ($p<0.001$). Bayesian model comparison revealed that the model that best explained the data is model S5i, a three-level network composed of bilateral A1 and STG and right IFG (see Fig. 4b). See Fig. 3 for the prior locations on the nodes of the network. For the winning model S5i, a *post hoc t*-test confirmed a significant coupling decrease for deviants *vs.* standards ($p<0.003$) in the backward connection linking rIFG to rSTG, and a trend increase ($p<0.1$) for the intrinsic connection within rA1 and the forward connection linking lA1 to lSTG.

Having identified the most likely network, we then finessed our search of model space by investigating where plasticity was most likely to be expressed; within the network architecture of winning model S5i. Six models were tested, encoding the hypotheses that differences in evoked responses (deviants *vs.* standards) were caused by connectivity changes in forward connections (F-model); changes in conjoint forward and intrinsic connections (Fi-model); changes in backward connections (B-model); conjoint backward and intrinsic connections (Bi-model); conjoint changes in forward and backward connections (FB-model); and changes in forward, backward, and intrinsic connections (FBi-model) (see Fig. 5a for details of model specification). Model FBi is identical to model S5i, the winning model in the first model search (see Fig. 4b). As expected, and in agreement with previous findings (Garrido et al., under review) the winning model was FBi (see Fig. 5b). We performed a repeated measures ANOVA of the log-evidences of the six models to assess the evidence for different model attributes, in relation to between-subject variability. The ANOVA had two factors, intrinsic connectivity (absent or present) and extrinsic connectivity (forward, backward, or both). This test revealed a trend effect of extrinsic connectivity ($p<0.1$) and a significant effect of intrinsic connectivity ($p<0.02$). Note that this does not change the fact that the best model, in terms of explaining all the data analysed, was the FBi-model; rather, it shows there is evidence for intrinsic adaptation under all the architectures we considered.

Discussion

Using dynamic causal modelling, we have presented further evidence to suggest: (i) the mismatch negativity (MMN) is generated by self-organised interactions within a hierarchy of cortical sources (Garrido et al., 2007) and (ii) the MMN rests on plastic change in both extrinsic (between-source) and intrinsic (within source) connections (Garrido et al.; under review). Critically, these conclusions are consistent with previous analysis of a conventional oddball paradigm but can now be generalised to the roving paradigm and indeed the notion of repetition–suppression in general (Desimone, 1996). Specifically, we investigated the outcome of stimulus repetition on scalp electroencephalographic responses and studied the underlying dynamics of the cortical network that generates these responses. Subjects were presented with a roving paradigm, a modified auditory oddball paradigm with standard tones that change sporadically to another frequency. Deviant tones elicited an MMN response peaking at about 180 ms over frontal channels (see Fig. 1b, c and Fig. 2). The difference wave between responses to deviants and responses to standards (here assumed to be established after the fifth repetition) revealed a statistical significant negativity over temporofrontal areas between 110 and 200 ms (Fig. 2b). This result is consistent with previous findings (Sams et al., 1985; Näätänen and Rinne, 2002; Baldeweg et al., 2004). Note that standards and deviants, as defined here, are physically identical; therefore, the MMN cannot be due to differential states of frequency-specific auditory neurons. The MMN could be explained by changes in the strength of the connectivity between and within the cortical sources of the underlying network. Changes in intrinsic connectivity are consistent with the idea that stimulus-specific adaptation (SSA) in A1 contributes to the emergence of the MMN (Ulanovsky et al., 2003). The decrease in inter-regional connection strengths over repetitions is consistent with predictive coding theories (Rao and Ballard, 1999; Friston 2005). From this perspective, perceptual learning of the auditory context may be understood as a process of (between-trial) prediction error suppression, implemented neurophysiologically through changes in connection strengths within a hierarchical cortical network (Friston 2005, Baldeweg 2006). See Jääskeläinen et al. (2007) for a discussion of the adaptation and short-term plasticity as driven by bottom-up and top-down effects.

It could be argued that we have exploited a false dialectic between the adaptation and model-adjustment hypotheses. The adaptation hypothesis pertains to neurophysiological mechanisms, whereas model-adjustment speaks to functional or perceptual mechanisms. It is likely that adaptation is an integral part of model-adjustment; in that adaptation mediated by changes in the coupling between remote sources may be involved in learning (see Winkler, 2007 and Jääskeläinen et al., 2007). The main conclusion from this study is that learning and implicit model-adjustments induce changes in both local and extrinsic coupling. This may reconcile physiological and functionalist explanations; furthermore, it highlights the utility of physiologically constrained models of functional architectures to explain data.

Technical issues

A feature of DCM, and hypothesis-driven methods in general, is that one cannot test all possibilities; *i.e.*, one has to constrain the model space in order to reduce it to a limited number of testable hypotheses. The search of model space does not offer an exhaustive exploration and selection of models; it only selects the best model amongst the models considered. Hence, there might be better models that explain connectivity changes during stimulus repetition or learning. This is a question of model comparison, and provided that there is good motivation for adding another model to the space of models, one can use Bayesian model comparison to evaluate a new model. An important consideration, when comparing DCMs, is that the free-energy accounts for both model accuracy and complexity. Therefore, it allows for comparison of models with different numbers of parameters (Friston et al., 2006). In brief, the free-energy can be decomposed into an accuracy term and a complexity term. The complexity term (the divergence between the prior and conditional density on the parameters) penalises models with a greater number of parameters and prevents selection of models that over-fit or do not generalise (see Penny et al., 2004 and Friston et al., 2006 for details).

DCM uses a conventional formulation of source localization (*c.f.* ECD, see Kiebel et al., 2006), but represents a departure from conventional source reconstruction or inverse solutions to the EEG problem by using a full spatiotemporal forward model that places constraints on the way sources activity is generated. Put simply, these constraints are that neuronal activity in one part of the brain must be caused by activity in another (David et al., 2006). Conventional methods localise sources associated with a specific peak and latency. In contrast, DCM explains a whole time-window, in this paper the 0 to 250 ms. Therefore, the models considered here attempt to explain the dynamics during the whole time interval.

As mentioned above, classical ECD solutions are part of DCM inversion. Here we used priors from previous studies for their mean locations and a variance of 16 mm^2; the orientations were estimated under uninformative or flat priors. The reason we used tighter priors on the location than on the orientations (moments) was that there is relatively little information in the EEG measurements about the spatial location of sources (therefore, changing the location priors would not change the results very much). In contrast, there is an enormous amount of information about their orientation. Evaluations of DCM with somatosensory evoked potentials revealed that precision on the orientation is substantially greater than the precision on location (Kiebel et al., 2006). Informative location priors can be derived from conventional source reconstruction techniques (David et al., 2006), classical ECD procedures, fMRI analyses, or from the literature, as used here.

Mechanisms of MMN generation

Mechanistic accounts of MMN generation posit changes in plasticity in extrinsic, forward and backward connections, between and hierarchical cortical sources (Friston, 2005). In this context, the difference waveform (and the MMN) arises from changes in coupling within and among cortical sources. We have previously used DCM to explain ERPs to standards and deviants and tested different plausible mechanisms or generative models for the MMN (Garrido et al., under review). For an internal consistency, the same models were tested here, given new data and a different paradigm. The choice of models was based on previous theoretical formulations of MMN generation, specifically *adaptation* (Jääskeläinen et al., 2004), *model-adjustment* (Winkler et al., 1996), and conjugations of two, which are in line with predictive coding. The predictive coding framework encompasses the two distinct hypotheses, in the sense that it predicts the adjustment of a generative model of current stimulus trains (*c.f. the model-adjustment hypothesis*) combined with local changes in post-synaptic sensitivity (*c.f. the adaptation hypothesis*). In agreement with our previous study, the best model comprised five reciprocally connected sources (bilateral A1 and STG, and right IFG). This is an important finding because it offers an comprehensive framework to explain the MMN; and furnishes direct evidence that the MMN is caused by self-organised changes in a cortical network with multiple hierarchical levels.

The MMN, a marker for auditory perceptual learning

The MMN reflects error detection caused by an unexpected or unlearned event that follows the perceptual learning of standards. This has been formulated under predictive coding (Friston 2003, 2005; Garrido et al., 2007); *i.e*., experience-dependent plasticity might underlie the perceptual discrimination of sounds (standards and deviants). These ideas are rooted in predictive coding models based on hierarchical Bayes (Rao and Ballard, 1999). In this framework, evoked responses correspond to prediction error that is explained away (within-trial) by neuronal dynamics during perception and is suppressed (between trials) by changes in connectivity during learning. Therefore the MMN can be interpreted as a failure to suppress prediction error, which can be explained quantitatively in terms of coupling changes among cortical regions. The repeated presentation of tones leads to learning or establishing a representation of a standard. This may render suppression of prediction error more efficient, leading to a reduction in evoked responses and the emergence of a mismatch response, when novel and therefore unlearned stimuli are presented. The suppression of evoked responses, due to a repeated event, is a 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). Changes in intrinsic connectivity would be caused by an initial adaptation phenomenon in the auditory cortices to repeated sounds and subsequent change detection when a different event, with different physical properties, occurs. From an empirical Bayesian perspective (*c.f.*, predictive coding), modulations in the intrinsic connectivity may encode changes in the precision of top-down predictions, responsible for suppressing prediction error. Changes in forward connections may reflect changes in 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 MMN represents a failure to predict bottom-up input and consequently a failure to suppress prediction error, which can be explained quantitatively in terms of coupling changes among and within cortical regions.

Unlike the traditional oddball paradigm, the roving-standard paradigm offers the possibility to follow the process of a deviant sound turning into a standard. This learning process involves dynamic connectivity changes that can be tracked with DCM. This will be the focus of future papers.

Acknowledgments

The Wellcome Trust and the Portuguese Foundation for Science and Technology funded this work. We thank David Bradbury for technical support and the volunteers for participating in this study.

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

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2008.05.018.

References

Baldeweg, T., 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends Cogn. Sci. 10, 93–94.

Baldeweg, T., Klugman, A., Gruzelier, J., Hirsch, S.R., 2004. Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr. Res. 69, 203–217.

Cowan, N., Winkler, I., Teder, W., Näätänen, R., 1993. Memory pre-requisites of mismatch negativity in the auditory even-related potential (ERP). J. Exp. Psychol. Learn. Mem. Cogn. 19, 909–921.

David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755.

David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. NeuroImage 25, 756–770.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG and MEG. NeuroImage 30, 1255–1272.

Desimone, R., 1996. Neural mechanisms for visual memory and their role in attention. Proc. Natl. Acad. Sci. U. S. A. 93, 13494–13499.

Doeller, C.F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., Schröger, E., 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. NeuroImage 20, 1270–1282.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530.

Friston, K., 2003. Learning and inference in the brain. Neural Netw. 16, 1325–1352.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond., B Biol. Sci. 360, 815–836.

Friston, K.J., Frith, C.D., Passingham, R.E., Liddle, P.F., Frackowiak, R.S., 1992. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. Proc. Biol. Sci. 248, 223–228.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2006. Variational free energy and the Laplace approximation. NeuroImage 34, 220–234.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36, 571–580.

Grau, C., Fuentemilla, L.I., Marco-Pallares, J., 2007. Functional neural dynamics underlying auditory event-related N1 and N1 suppression response. NeuroImage 36, 522–531.

Haenschel, C., Vernon, D.J., Prabuddh, D., Gruzelier, J.H., Baldeweg, T., 2005. Event-related brain potential correlates of human auditory sensory memory-trace formation. J. Neurosci. 25, 10494–10501.

Henson, R.N., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., Rugg, M.D., 2003. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cereb. Cortex 13, 793–805.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. U. S. A. 101, 6809–6814.

Jääskeläinen, I.P., Ahveninen, J., Belliveau, J.W., Raij, T., Sams, M., 2007. Short-term plasticity in auditory cognition. Trends Neurosci. 30, 653–661.

Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. 73, 357–366.

Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. NeuroImage 30, 1273–1284.

Kiebel, S.J., Garrido, M.I., Friston, K.J., 2007. Dynamic causal modelling of evoked responses: the role of intrinsic connections. NeuroImage 36, 332–345.

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G., Näätänen, R., 1999. Frequency change detection in human auditory cortex. J. Comput. Neurosci. 6, 99–120.

Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event related potentials and other brain measures of cognitive function. Behav. Brain Sci. 13, 201–288.

Näätänen, R., 1992. Attention and brain function. Laawrence Erlbaum, Hillsdale, New Jersey.

Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol. Bull. 125, 826–859.

Näätänen, R., Rinne, T., 2002. Electric brain response to sound repetition in humans: an index of long-term-memory — trace formation? Neurosci. Lett. 318, 49–51.

Näätänen, R., Jacobsen, T., Winkler, I., 2005. Memory-based or afferent process in mismatch negativity (MMN): a review of the evidence. Psychophysiology 42, 25–32.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., Schröger, E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. NeuroImage 15, 167–174.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. NeuroImage 22, 1157–1172.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.-J., Zilles, K., 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. NeuroImage 13, 669–683.

Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., Näätänen, R., 2000. Separate time behaviors of the temporal and frontal mismatch negativity sources. NeuroImage 12, 14–19.

Rockland, K.S., Pandya, D.N., 1979. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res. 179, 3–20.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R., 1985. Auditory frequency discrimination and event-related potentials. Electroencephalogr. Clin. Neurophysiol. 62, 437–448.

Sams, M., Alho, K., Näätänen, R., 1993. Sequential effects on the ERP in discriminating to stimuli. Biol. Psychol. 17, 41–58.

Sussman, E., Winkler, I., 2001. Dynamic sensory updating in the auditory system. Brain Res. Cogn. Brain Res. 12, 431–439.

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391–398.

Wager, T.D., Keller, M.C., Lacey, S.C., Jonides, J., 2005. Increased sensitivity in neuroimaging analysis using robust regression. NeuroImage 26, 99–113.

Winkler, I., 2007. Interpreting the mismatch negativity (MMN). J. Psychophysiol. 21, 147–163.

Winkler, I., Karmos, G., Näätänen, R., 1996. Adaptive modelling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res. 742, 239–252.