Predictive Coding and Pitch Processing in the Auditory Cortex

Sukhbinder Kumar[1,2], William Sedley[1], Kirill V. Nourski[3],
Hiroto Kawasaki[3], Hiroyuki Oya[3], Roy D. Patterson[4],
Matthew A. Howard III[3], Karl J. Friston[2],
and Timothy D. Griffiths[1,2]

Abstract

■ In this work, we show that electrophysiological responses during pitch perception are best explained by distributed activity in a hierarchy of cortical sources and, crucially, that the effective connectivity between these sources is modulated with pitch strength. Local field potentials were recorded in two subjects from primary auditory cortex and adjacent auditory cortical areas along the axis of Heschl's gyrus (HG) while they listened to stimuli of varying pitch strength. Dynamic causal modeling was used to compare system architectures that might explain the recorded activity. The data show that representation of pitch requires an interaction between nonprimary and primary auditory cortex along HG that is consistent with the principle of predictive coding. ■

INTRODUCTION

Mechanisms for pitch perception are a subject of controversy, with some studies suggesting the existence of single areas (Bendor & Wang, 2005; Penagos, Melcher, & Oxenham, 2004; Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lutkenhoner, 2003) and others suggesting distributed processing over areas (Griffiths et al., 2010; Bizley, Walker, Silverman, King, & Schnupp, 2009). We consider here the idea that pitch perception requires a functional *system* comprising several areas with specific patterns of effective connectivity between them. We test this idea by comparing different dynamic causal models of electrical activity recorded directly from human auditory cortex using depth electrodes: We were particularly interested in testing biophysical models with a hierarchical connectivity based on a predictive coding account of pitch perception.

From a psychophysical perspective, pitch is a fundamental auditory percept with a complex relationship to the structure of the sound in frequency and time (see de Cheveigné, 2005, for a review). From a biological perspective, this suggests that the representation of pitch by the brain will not rest on a simple mapping of stimulus properties such as frequency. The auditory cortex of mammals contains multiple areas, each containing systematic frequency mappings, with mirror reversal of frequency gradients between areas (Kaas & Hackett, 2000). Recordings from single neurons have looked at whether some of these areas might be specialized for the representation of pitch. In the marmoset, neurons that show a form of "pitch tuning" have been demonstrated in a low-frequency area abutting primary cortex in A1 (Bendor & Wang, 2005), whereas in the ferret, selective responses to pitch (based on a less strict criterion for pitch responsiveness) have been demonstrated in multiple areas (Bizley et al., 2009).

In humans, direct recordings of local field potentials (LFPs) show responses to temporally regular sounds when these have rates associated with pitch (Griffiths et al., 2010). The responses are found in human primary cortex in medial Heschl's gyrus (HG) and adjacent nonprimary areas in HG. fMRI studies (Puschmann, Uppenkamp, Kollmeier, & Thiel, 2010; Penagos et al., 2004; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002) demonstrate maximal activity in lateral HG activity during pitch perception, although activity does occur in more medial areas too (see Griffiths et al., 2010, for discussion). Megneto-encephalography (MEG) studies (Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004; Krumbholz et al., 2003; Gutschalk, Patterson, Rupp, Uppenkamp, & Scherg, 2002) have also demonstrated activity that is lateral to primary auditory cortex. These studies beg the question as to how activity in the primary auditory cortex in medial HG and nonprimary auditory cortex in more lateral parts of HG is related.

Predictive coding (Friston & Kiebel, 2009; Friston, 2002a,b, 2005; Rao & Ballard, 1999; Mumford, 1992) as a model for perception posits that the brain uses a hierarchical generative model to predict and explain sensations. Representations of the causes of sensory input (e.g., temporal regularity for pitch perception) are optimized by

[1]Newcastle University Medical School, [2]University College London, [3]University of Iowa Hospitals and Clinics, [4]University of Cambridge

minimizing prediction error: Predictions are passed to lower levels of a cortical sensory hierarchy by backward connections where they are compared with low-order representations (or sensory input at the lowest level) to produce a prediction error. The prediction error is then sent back to the level above via forward connections to improve the predictions, hence reduce prediction error. This iterative process continues, until the prediction error is minimized and an optimal hierarchical representation is formed. This model forms a theoretical basis for both visual (Kersten, Mammasian, & Yuille, 2004; Rao & Ballard, 1999) and auditory (Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009) perception. We hoped to find evidence for this hierarchical message-passing by comparing different (hierarchical and nonhierarchical) connectivity models of observed electrophysiological responses.

In the present study, LFPs were recorded from primary auditory cortex and adjacent auditory cortical areas along the axis of HG while subjects listened to stimuli with varying pitch strength. We examined effective connectivity using dynamic causal modeling (DCM; David et al., 2006) and Bayesian model selection (Penny et al., 2010; Stephan, Penny, Duanizeau, Moran, & Friston, 2009) to determine (i) the effective connectivity between medial, middle, and lateral HG and (ii) how these connections are modulated with varying pitch strength.

In addition to quantifying effective connectivity between areas, DCM allows the comparison of hierarchical architectures within the auditory system by defining forward connections (from lower to higher areas), parallel connections (between areas at the same level in the hierarchy), and backward connections (from higher to lower areas). In our DCM, forward connections are modeled as originating in pyramidal cells and targeting granular layers, whereas backward connections target supragranular and infragranular layers (cf. Felleman & Van Essen, 1991). We hypothesized (i) that lateral HG is at a higher level in the auditory hierarchy than medial HG and (ii) that the top–down influence of higher areas (lateral HG) would increase with the predictability (strength) of pitch, in accord with the predictive coding model; that is, backward connections would predominate over forward connections. Our results demonstrate prominent effective connectivity between the three areas consistent with a hierarchical architecture and pitch strength-dependent changes in effective connectivity between lateral HG and lower areas that are consistent with predictive coding.

METHODS

Dynamic Causal Modeling: Theory

In conventional noninvasive studies of brain function, brain responses using EEG and MEG or fMRI are routinely measured in response to a stimulus or when a cognitive/motor task is performed. However, most of the interesting things that happen when the brain is activated are hidden (that is, not directly measurable). For example, activity measured at one site of the brain may not be the sole result of processing at that site, but it may also reflect neuronal interactions between areas. The goal of DCM is to make inferences about the hidden parameters and variables using measured variables. Specifically, DCM tries to explain the observed brain responses in terms of underlying causal interactions between different areas at the neuronal level. The technique was first used to infer the neuronal interactions from the measured BOLD signals from fMRI (Friston, Harrison, & Penny, 2003). Subsequently, it has been extended to EEG/MEG (David et al., 2006) and LFPs (Moran et al., 2009). Here we apply DCM to LFPs recorded directly from human auditory cortex. DCM has four components: (i) specification of a biologically realistic neuronal model for each area, (ii) specification of models of causal interactions or extrinsic coupling among different areas, (iii) selection of the best model or architecture based on the evidence in the data, and (iv) inference of the parameters of the best model, given those data.

A single source in DCM is modeled by "a neural mass model." The idea behind the neural mass model is that the state of an ensemble of neurons at a given time can be characterized by the mean activity of the ensemble. The dynamics of an ensemble over time can, therefore, be characterized by how this mean activity evolves over time and can be specified formally with biologically constrained differential equations (see Deco, Jirsa, Robinson, Breakspear, & Friston, 2008, for a review of neural mass models). The neural mass model used in DCM was first described by Jansen and Rit (1995) and comprises three populations of neurons: A population of (excitatory) pyramidal cells receives inputs from excitatory and inhibitory interneurons (Supplementary Figure 1A). In DCM, each source is modeled with a three-population Jansen and Rit model, where the subpopulations are assigned to three layers: supragranular, granular, and infragranular layers. Supragranular and infragranular layers comprise the superficial and deep pyramidal cells, respectively, along with a population of inhibitory interneurons. The granular layer consists of excitatory interneurons (cf. spiny stellate cells) only (Supplementary Figure 2A). Synaptic dynamics are modeled as a linear system, which is characterized by a (postsynaptic response) kernel with two parameters for each subpopulation: a time constant and a maximum amplitude. Presynaptic activity is convolved with the kernel to produce postsynaptic activity. This is transformed by a nonlinear sigmoid function to firing rate (see David et al., 2006; Jansen and Rit, 1995, for details). The output measured at a given area is modeled as a mixture of depolarization of each of the three populations (that is dominated by contributions from the pyramidal cells).

Cortico-cortical connections between different areas are arranged hierarchically. This hierarchy is reflected in the laminar pattern of origin and termination of connections between the two areas (Felleman & Van Essen, 1991). Specifically, forward connections originate in the supragranular

layers and terminate in the granular layer, whereas backward connections originate in the infragranular layers and terminate in agranular layers and lateral connections connect agranular layers (Supplementary Figure 2B). This means that different areas can be connected by extrinsic connections that follow these anatomical rules. Each pattern of connections represents a different hypothesis about the functional architecture and corresponds to a competing model or DCM. Implementation of these connections using the Jansen and Rit (1995) model is shown in Supplementary Figure 1B.

The final stage of DCM is the selection and optimization of their parameters using measured brain responses. Mathematically, any DCM can be described by two equations:

$$\frac{dx}{dt} = f(x, u, \theta)$$

$$y = g(x, \theta). \tag{1}$$

The first (state) equation specifies how the experimental input $u(t)$ influences the dynamics of hidden states $x(t)$ and the second (observer) equation links the hidden states $x(t)$ to measured brain responses $y(t)$. θ represents the unknown parameters of model-like connection strengths and synaptic parameters, which are to be estimated. The parameters are estimated using Bayesian statistics, which specify the posterior density of parameters θ, given the data:

$$p(\theta|y, m) = \frac{p(y|\theta, m)p(\theta|m)}{p(y|m)}. \tag{2}$$

Where $p(y|\theta, m)$ and $p(\theta|m)$ are the likelihood and prior density of parameters θ, respectively, of a given model m. The denominator $p(y|m)$ is called the model evidence and is calculated as

$$p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta. \tag{3}$$

An iterative method called variational Bayes (Friston, 2002a, b) is used to estimate the posterior density $p(\theta|y, m)$ and the model evidence $p(y|m)$. In this method, posterior density is approximated by density $q(\theta)$ that is assumed to be Gaussian. The idea behind variational Bayes is that the model evidence can be expressed as

$$\ln p(y|m) = F(q, \theta, m) + D(q \| p(\theta|y, m)) \tag{4}$$

or

$$F(q, \theta, m) = \ln p(y|m) - D(q \| p(\theta|y, m)), \tag{5}$$

where F is the free energy and $D(q \| p(\theta|y, m))$ is Kullback–Leibler distance between density q and posterior density $p(\theta|y, m)$. Because Kullback–Leibler distance is nonnega-tive, maximization of free energy minimizes the distance between q and posterior density. That is, q approximates the posterior density: $q \rightarrow p(\theta|y, m)$. Furthermore, the maximum value of free energy approximates model log evidence, that is,

$$\ln p(y|m) \approx F(q, \theta, m). \tag{6}$$

The log evidence for different models can be used to determine the best model, given some data. DCM is usually used as a hypothesis-driven technique, where a number of models or hypotheses are specified in advance, and the log evidence for each model is calculated using the free energy approximation above. A complete list of parameters θ that are optimized is given in David et al. (2006).

Subjects, Surgery, and Recording

LFPs were recorded from two adult subjects, R154 and L156, undergoing intracranial electrophysiological recording to localize epileptic foci. Both subjects had normal hearing as confirmed by audiometric testing before implantation of electrodes. Hybrid depth electrodes (Reddy et al., 2010; Howard et al., 1996) with 14 high impedance contacts (70–300 kΩ) were implanted along the long axis of HG in right hemisphere for subject R154 and left hemisphere for subject L156. The electrode contact positions were determined by coregistering electrode locations identified on postoperative MRI scans with the subject's preoperative three-dimensional brain MRI. The localization procedure demonstrated that all experimental high impedance electrode contacts in subject R154 and all, but contacts 13 and 14 in subject L156 were in gray matter, along the axis of HG. The research protocols were approved by the University of Iowa Human Subjects Review Board. Prior informed consent was obtained from each subject before the study. Figure 1 shows the electrode locations in the two subjects.

Electrical activity and effective connectivity were examined for three contacts in the medial, middle, and lateral part of HG in each subject. For subject R154, the selected representative contacts were 1, 8, and 14; for subject L156, the contacts were 1, 7, and 12. The corresponding Talairach coordinates for these electrodes (Supplementary Tables 1 and 2) show that they are located at the three sites of maximal activity for sound minus silence contrasts in fMRI (Patterson et al., 2002), where the medial site corresponds to primary auditory cortex (human homologue of A1). The lateral maxima may correspond to homologues of nonprimary areas in macaque (Brugge et al., 2009; Hackett, 2007).

Stimuli

The stimuli consisted of a 1-sec burst of broadband noise followed by 1.5 sec of regular interval noise (RIN). The RIN was created using a delay-and-add algorithm (Yost, 1996).
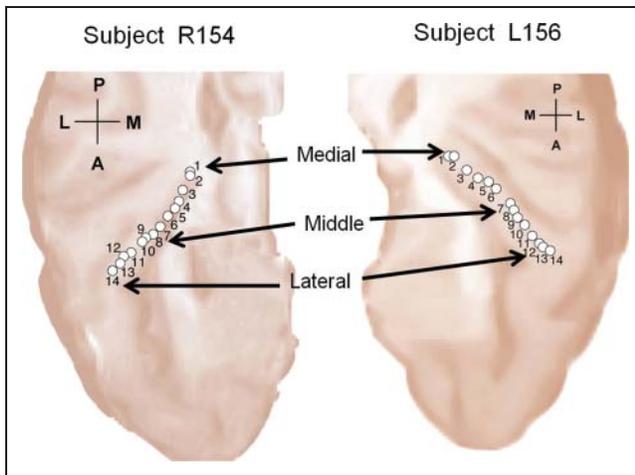
Figure 1. Electrode locations for the two subjects (subject R154 and L156) along the axis of HG overlaid on the MRI of the superior temporal plane. Three contacts, one each in the medial, middle, and lateral parts of HG, were considered in the effective connectivity analyses. For subject R154, the chosen contacts were 1, 8, and 14 and for subject L156 the contacts were 1, 7, and 12.

RIN is also known as "iterated rippled noise" because of the ripples that the delay-and-add process induces in the frequency magnitude spectrum of the stimulus. We use the term RIN here to emphasize the temporal cue observed in the pattern of neural firing in the auditory nerve and the temporal cue used in models of RIN perception (Patterson, Handel, Yost, & Datta, 1996; Yost, Patterson, & Sheft, 1996). The delay in the delay-and-add cycle determines the pitch value that the listener hears, and the number of cycles or iterations determines the pitch strength or salience (Patterson et al., 1996; Yost et al., 1996). The stimuli were normalized to a common power spectral density, high-pass filtered using a cutoff frequency of 800 Hz (to remove spectral ripples that might be resolved by the cochlea) and masked with broadband noise below the cutoff frequency (Griffiths et al., 2010).

Paradigm

Recordings were made in a dedicated recording facility in a shielded room. The experiments employed a passive listening paradigm. Subjects were awake with eyes open and relaxed during the recording sessions. The stimuli were delivered diotically via Etymotic ER4B earphones in custom earmolds at a comfortable sensation level of 45–55 dB. The DCM analysis was based on data acquired with RIN constructed with 8, 16, and 32 iterations and a fixed pitch value of 128 Hz. There was also a baseline condition with zero iterations, that is, a spectrally matched noise with no pitch. Time series were recorded from each electrode and averaged over 50 repetitions for each stimulus condition.

Data Preparation

LFPs were down sampled to 250 Hz, band-pass filtered between 4 and 16 Hz, and averaged across trials. This narrow range of frequency band was chosen to analyze only the evoked responses time locked to stimulus onset. Evoked responses during the first 300 msec after RIN onset were analyzed.

DCM Specification

The principle objective of the present analysis was to ask, (i) what types of connections (forward, backward or lateral) couple the medial, middle, and lateral areas of HG and (ii) how are these connections modulated during the processing of stimuli with increasing pitch strength? To address the first question, a model space (set of models) was constructed based on the following biologically informed criteria:

- If an area A sends forward connections to area B, then B sends backward connections to area A.
- If an area A sends lateral connections to area B, then B sends lateral connections to area A.

Because there are six connections among the three areas and the three are fixed by the above constraints (e.g., if the connection between Regions A and B is specified as forward, then it follows that connection from B to A is backward), there are three unspecified connections, each of which could be forward, backward, or lateral). This gives $3^3 = 27$ possible models.

To address the second question, a model space was constructed in which every connection in the model is either modulated or not modulated by temporal regularity. Because there are six connections, there are $2^6 = 64$ models for each combination of connection types.
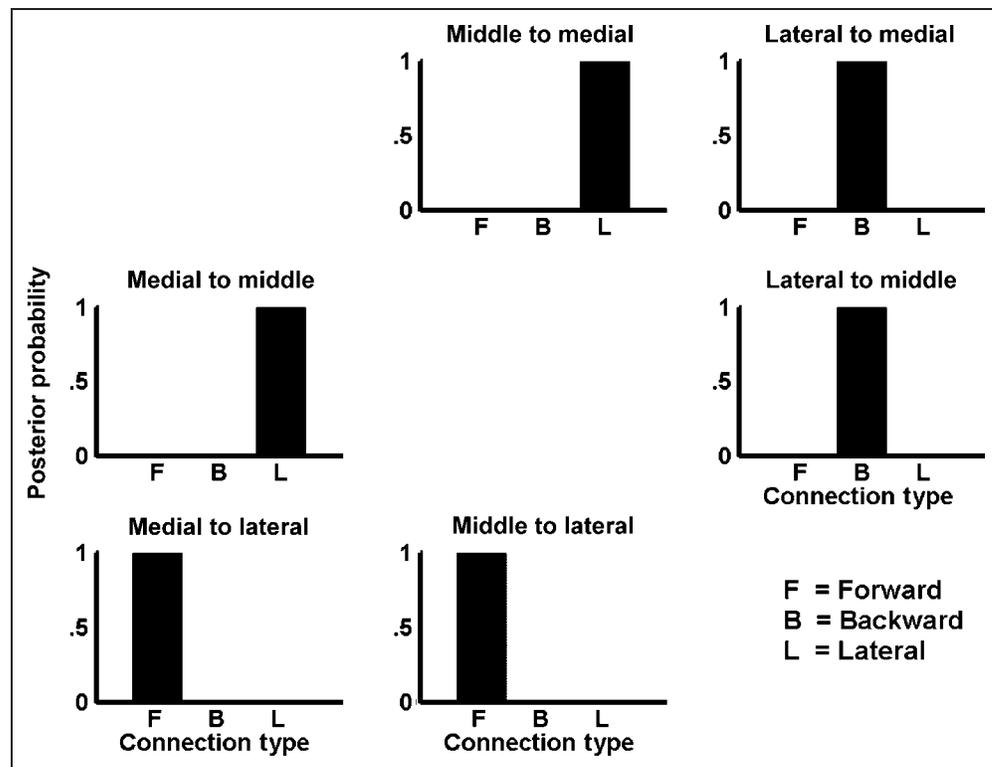
To finesse an exhaustive search over (27×64) models with different connections and modulations, we used a heuristic search strategy in which we first optimized the connection types (over subjects and pitch strength) and then optimizing modulation-type models with the ensuing connection types.

The DCMs exogenous inputs ($u(t)$ in Equation 1 above) comprise input relayed by subcortical structures and were modeled by gamma functions (David et al., 2006). In the present study, we used four gamma functions (Supplementary Figure 3). A gamma function models the event-related input that is delayed with respect to stimulus onset (this parameterization of the inputs was optimized using Bayesian model comparison, with one to six gamma functions). This exogenous input entered all three regions of HG. We used multiple input components (gamma functions) to model the unknown convolutions of sensory discharges by earlier (subcortical) systems.

Family-wise Model Comparison

Because the connection types among medial, middle, and lateral regions are not constrained by the nature of the

Figure 2. Posterior probability of model families, where each family (or partition of model space) was defined in terms of the connection type for each connection. The posterior probability was computed using fixed effect analysis over three conditions (8, 16, and 32 iterations) and two subjects.

exogenous input, all 27 connection-type models were inverted for different levels of temporal regularity (8, 16, and 32 iterations) and both subjects. To determine the type of a given connection (e.g., between medial and middle regions), all the models (across all regularity and subjects) were divided into three families: Family F1, in which the connection was forward; family F2, in which the connection was backward; and family F3, in which the connection was lateral. The posterior probability that each connection was forward, backward, or lateral was computed by summing the posterior probabilities of all (nine) models in each family. The posterior probability of each model was evaluated by summing the log evidence for each of the (27) models over subjects and regularity (under the assumption of independent data from each observation). The exponential of these pooled log evidences was normalized so that their sum was unity. This gives the posterior model probability, under prior assumptions that each model was equally probable. Having established the optimum connection types, we then inverted all 64 modulation-type models and examined the best models to see how temporal regularity (pitch strength) modulated those connections.

RESULTS

Type of Connections between Medial, Middle, and Lateral Regions of HG

We constructed a model space consisting of 27 models that spanned all possible hypotheses about the types of connections between medial, middle, and lateral regions

of HG. The family-wise posterior probabilities for each connection being forward, backward, or lateral are shown in Figure 2. This figure shows that medial and middle regions are connected to each other by lateral connections, whereas the lateral part of HG receives forward connections from and sends backward connections to both the medial and middle part of HG. A schematic representation of this architecture is shown in Figure 3. On the basis of the hierarchal specificity of laminar projections (Felleman & Van Essen, 1991; Maunsell & Van Essen, 1983), these results suggest as follows:

• The medial and middle parts of HG are reciprocally connected by lateral connections and are at a similar level of hierarchy.
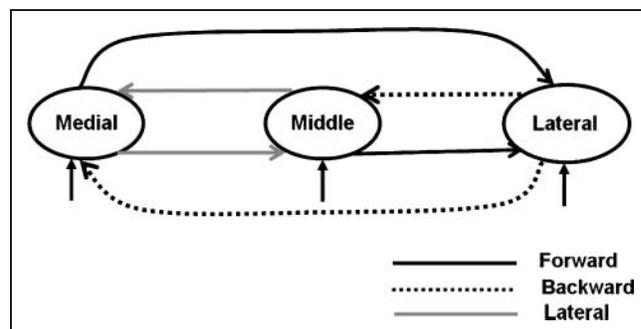• The lateral part of HG is at a higher level of the auditory hierarchy than medial and middle parts.

Figure 3. Most probable connection types between medial, middle, and lateral parts of HG.

Modulation of Connectivity by Temporal Regularity

Having established the types of connection, we next investigated how these connections were modulated by the temporal regularity of the RIN. Event-related responses to RIN with 0, 8, 16, and 32 iterations from the medial, middle, and lateral HG were analyzed together in a single DCM. This involved optimizing additional parameters that controlled how pitch strength (number of RIN iterations) modulated the strength of connections monotonically, over the four ERPs (as in Garrido et al., 2008). We constructed 64 variants of the model shown in Figure 3. These models were based on all possible combinations of how pitch strength could modulate extrinsic connections among the three areas. Posterior probabilities for each of these 64 models for the two subjects R154 and L156 are shown in Figure 4A and B, respectively. For subject R154, there are two comparably plausible models (64 and 48) that have posterior probabilities of .52 and .37, respectively. For subject L156, the best model (model 60) has a posterior probability of .78 and the second best model (model 44) has a posterior probability of .20. The best models (64 and 48 for subject R154 and 60 and 44 for subject L156) for the two subjects are shown in Figure 5. Red and green triangles denote those connections that are modulated by pitch strength. These results show that in subject R154 (Figure 5A), the two winning models have a very similar structure: In model 64 (posterior probability = .52), all the connections are modulated, whereas in model 48 (posterior probability = .37), all but the middle to medial connection are modulated by temporal regularity. In subject L156 (Figure 5B), the best model (model 60, posterior probability = .78) requires modulation of all connections with the exception of lateral to middle whereas in the second best model (model 44, posterior probability = .2), in addition to the connection in the best model, the connection from middle to medial is also not modulated.
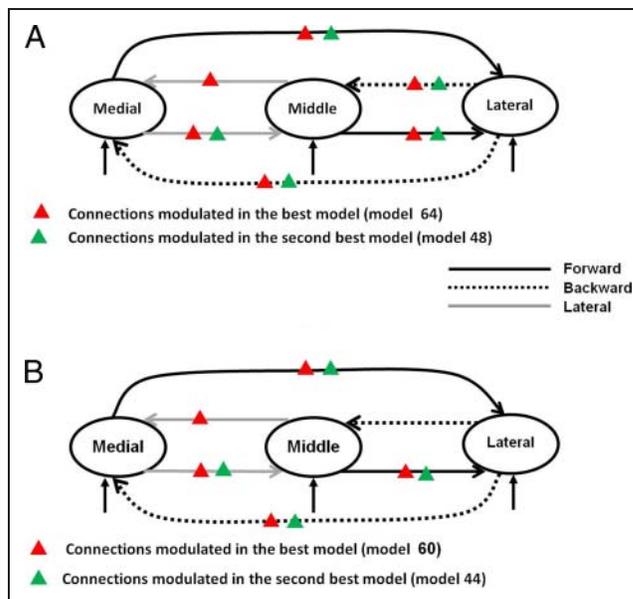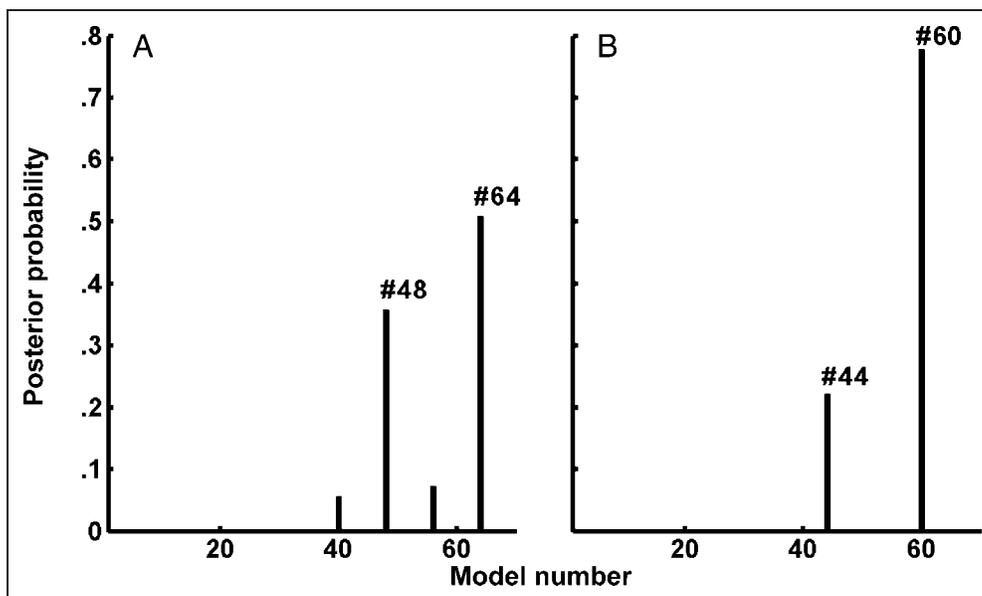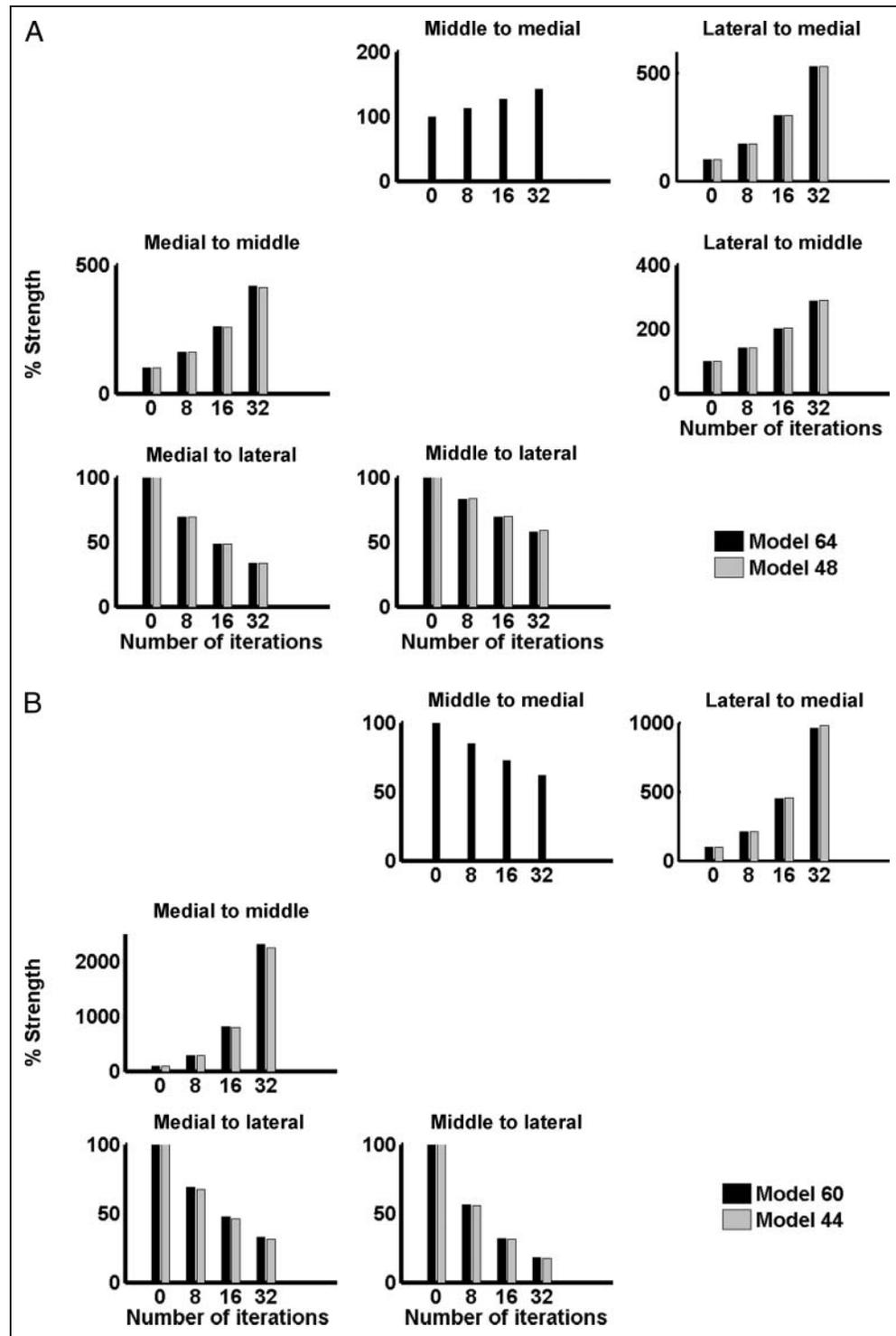

Figure 5. Structure of the best models for subject R154 (A) and for subject L156 (B).

Figure 6 plots the change in connection strength with temporal regularity for both subjects. Modulation of connectivity for the best model (model 64 in subject R154, Figure 6A, and model 60 in subject L156, Figure 6B) is shown in black. Modulation for the second best model (model 48 in subject R154 and model 44 in L156) is shown in gray. The profile of modulation is remarkably consistent between the two subjects and shows distinct effects of pitch strength on different connections within the system. The following generalizations can be drawn from these results:

- For subject R154, all connections show very similar patterns of pitch strength modulation, except the

Figure 4. Posterior probabilities of 64 modulation-type models for subject R154 (A) and for subject L156 (B).

Figure 6. Modulation of connectivity with temporal regularity for subject R154 (A) and for subject L156 (B).

connection from middle to medial region, which is modulated in one model (model 64) but not the other (model 48).

- For subject L156, the pattern of connectivity is again very similar except in the middle to medial region, which is modulated in the best model (model 60) but not in the second best model (model 44).
- Backward connections from lateral HG (to both medial and middle HG in subject R154 and to only medial HG in

subject L156) increase with temporal regularity. In both subjects, there is almost a doubling of connection strength with increasing temporal regularity.

- Forward connections from both medial and middle HG decrease with temporal regularity.
- Lateral connection strengths (from medial to middle and middle to medial) increase with temporal regularity. However, the medial to middle connection changes much more than the reciprocal connection.

DISCUSSION

Connection Types in HG

On the basis of cytoarchitectonics, Brodmann (1909) localized primary auditory cortex to HG. However, further studies have shown that HG is not a single homogeneous area but consists of at least two areas (von Economo and Koskinas' (1925) areas TC and TD and Galaburda and Snides' (1980) KAm and kAlt) or three areas (Morosan et al.'s (2001) Te 1.0, Te 1.1, and Te 1.2). To the best of our knowledge, however, there is no literature on the types of connections that exist between these distinct regions in humans. We applied DCM to depth electrode data recorded from the medial, middle, and lateral regions of HG to infer the types of (effective) connections between them. Our results suggest that medial and middle regions are connected by lateral connections, whereas the lateral region receives forward projections from and sends backward connections to the other two regions of HG. This implies that lateral HG is at a higher level of the auditory hierarchy than the medial and middle regions (Felleman and Van Essen, 1991) and medial and middle regions of HG occupy similar levels.

The notion that lateral HG is at a higher level of hierarchy than medial HG agrees with a number of previous studies. Cytoarchitectonic studies in humans (Morosan et al., 2001; Galaburda & Snides, 1980) have shown that lateral HG is less "primary-like" than medial and middle HG. von Economo and Koskinas (1925) described this area as a "transition zone between primary and nonprimary areas" (Morosan et al., 2001). Although the homology between the auditory areas of macaque and human is not well established, functional studies using the same stimuli in both humans and macaque have suggested that lateral HG may correspond to area R/RT (Baumann et al., unpublished observations) or may correspond to a belt area (Brugge et al., 2009) in macaques. Moreover, the possibility that medial and middle regions of HG are at a similar hierarchical level is consistent with functional studies, which shows that medial and middle HG have similar responses and may both lie in the core area (Brugge et al., 2009).

Modulation of Connectivity Strength with Increasing Temporal Regularity

We have shown that backward connections from lateral HG (to medial and middle HG) increase with temporal regularity and forward connections (from medial and middle HG) decrease with temporal regularity. These results can be explained by predictive coding (Friston & Kiebel, 2009; Friston, 2002a,b, 2005; Rao & Ballard, 1999; Barlow, 1994; Mumford, 1992). The idea behind predictive coding is that, in a hierarchically organized brain, areas higher in the hierarchy (here lateral HG) use a generative model of the world to make predictions of representations at lower levels. These predictions are passed to lower areas by means of backward connections (here medial and middle

HG). The difference between the actual representation at the lower area and the prediction is the prediction error. This is passed back to the higher area by means of forward connections to adjust the higher level representation: If the error is large, then the model of the world "stored" in higher-order area is not correct and needs updating. This recursive message passing entails an iterative process, which aims at minimizing prediction error at all levels in the hierarchy, to describe the causes of sensory input at multiple levels. Clearly, our use of RIN speaks directly to the predictability of stimuli and the perceptual inferences about pitch. Our hypothesis assumed that as the predictability of stimuli increased the top–down influences mediating predictions would become stronger relative to bottom–up passing of prediction errors. The theoretical mechanism behind this effect is quite simple: In computational models of predictive coding the precision (inverse variance) of prediction error is encoded by the postsynaptic sensitivity of prediction error units, generally thought to be superficial pyramidal cells. This means that when stimuli are predictable (and prediction errors are low) the responsiveness of pyramidal cells to top–down predictions increases (because precision is high). This is what we observed empirically in the DCM. Similar results have also been found in studies of perceptual discrimination using endogenous fluctuations in activity or sensitivity (e.g., Hesselmann, Sadaghiani, Friston, & Kleinschmidt, 2010). This finding is also consistent with the relative decrease in the strength of forward connections for standard stimuli relative to unpredicted oddball stimuli using the DCM and MMN paradigm (e.g., Garrido et al., 2009).

The lateral connections between medial and middle HG are also modulated by temporal regularity. The medial to middle connection increased in both subjects and middle to medial connection increased for one subject and decreased for the other. One possible functional role of lateral connections is to decorrelate regions of the network, which respond to the same feature of the stimulus (Sirosh & Mikkulainen, 1996; Foldiak, 1990). For example, if two regions respond to the same stimulus feature, then responses of these two regions will be correlated. Lateral connections using mutual inhibition reduce this redundancy and make the representations more efficient (a sparse representation). This can be shown formally to be an emergent property of predictive coding (Friston, 2008).

There is a close relationship between the number of iterations used to generate a RIN and the strength of the pitch that listeners hear, and so the model of regularity processing has direct implications for models involving perceptual inferences about pitch and pitch strength. A number of previous fMRI studies (Penagos et al., 2004; Patterson et al., 2002) have emphasized a role for lateral HG in the processing of temporal regularity and the perception of pitch. Our results suggest a specific role for lateral HG in pitch prediction as part of a constructive (predictive) hierarchical model, distributed within an auditory pitch system.

A number of computational models for pitch perception have been proposed in the literature (de Cheveigné, 2005). Most of these models lack biological realism because (i) they are driven bottom–up: these models compute some feature (e.g., spectrum or autocorrelation of the stimulus without using any top–down information, and (ii) they are nonhierarchical: they extract the percept only at one scale. The current theories of brain function suggest that the percept is computed hierarchically at different time scales and is driven both by bottom–up and top–down flow of information (Friston, 2008). One such model (Balaguer-Ballester, Clark, Coath, Krumbholz, & Denham, 2009), emphasizing the role of hierarchies and top–down effects in computing pitch, was proposed recently. In this model, higher areas optimize the temporal scale over which information is integrated in lower areas. Thus, different temporal scales are invoked, depending on the (slow or fast) dynamics of the stimulus. We suggest that lateral HG may play a similar role and adapts the time scale of integration in lower areas (primary auditory cortex and subcortical areas) in a context-sensitive manner. This might be achieved using the prediction signal from lateral HG or the local prediction error signal (e.g., in primary auditory cortex) to adapt processing in primary auditory cortex. Please see Kiebel, von Kriegstein, Daunizeau, and Friston (2009) for a discussion of related mechanisms in tracking auditory sequences under the predictive coding framework.

The predictive coding hypothesis has number of consequences, some of which we have exploited when comparing different explanations (DCMs) for our data: These include (i) a hierarchy of cortical levels, (ii) forward and backward message-passing that entails reciprocal and directed connectivity, (iii) functional asymmetries in forward and backward connections (modeled here in terms of the subpopulations targeted), and (iv) top–down influences can only be expressed when predictions can be formed, suggesting a predictability-dependent (pitch salience-dependent) expression of backward effective connectivity.

Although not explicitly tested here, predictive coding also suggests (i) areas higher in the hierarchy (lateral HG in the present study) will have a longer temporal window of integration. This is because higher areas (which predict activity in lower areas) receive inputs (prediction errors) from a number of areas below, each of which integrates input using a smaller temporal window. (ii) The dynamics of areas higher in the hierarchy will unfold more slowly than areas lower in the hierarchy. (iii) Responses (context-sensitive predictions) to a given event depend on the context surrounding the event. For example, an MEG study (Chait, Poeppel, & Simon, 2007) showed that responses to transitions from an ordered train of tone pips to a disordered train are different when the transition is made in the reverse direction (that is, from ordered tone pips to disordered tone pips). See Friston (2008), for a fuller discussion of these issues.

One possible criticism of our study could be that we have used only one type of stimulus (RIN), and the analysis is restricted to areas lying along HG. A recent fMRI study (Hall & Plack, 2009) using a broader range of pitch-producing stimuli has shown that pitch-related activity may extend to areas beyond HG. However, the role of lateral HG in pitch perception is not restricted to RIN stimuli only. Studies from several groups using stimuli other than RIN have shown the role of lateral HG in pitch perception. These stimuli include harmonic complexes (Penagos et al., 2004; Warren, Uppenkamp, Patterson, & Griffiths, 2003), Huggins pitch (Puschmann et al., 2010), and click trains (Gutschalk et al., 2002, 2004). It will be interesting to see how specific the system we have identified is to the type of pitch used.

In our previous study (Griffiths et al., 2010), we observed both evoked and induced high-frequency gamma (80–120 Hz) in response to RIN all along HG. The latter particularly occurring when the RIN frequency was above the lower limit of frequency that is perceived as pitch. In the current study, we have only focussed on how to explain the evoked responses in terms of interactions between the medial, middle, and lateral part of HG. Interactions between these regions in the gamma range will be addressed in future studies using DCM for induced responses (Chen, Kiebel, & Friston, 2008).

REFERENCES

Balaguer-Ballester, E., Clark, N. R., Coath, M., Krumbholz, K., & Denham, S. L. (2009). Understanding pitch perception as a hierarchical process with top-down modulation. *PLoS Computational Biology, 5,* e1000301. doi:10.1371/journal. pcbi.1000301.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 1–22). Cambridge, MA: MIT Press.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in the primate auditory cortex. *Nature, 436,* 1161–1165.

Bizley, J. K., Walker, K. M., Silverman, B. W., King, A. J., & Schnupp, J. W. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *Journal of Neuroscience, 29,* 2064–2075.

Brodmann, K. (1909). *Vergleichende lokalisationslehre der grosshirnrinde*. Leipzig: Barth.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., et al. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology, 102,* 2358–2374.

Chait, M., Poeppel, D., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience, 98,* 224–231.

Chen, C. C., Kiebel, S. J., & Friston, K. J. (2008). Dynamic causal modeling of induced responses. *Neuroimage, 41,* 1293–1312.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., & Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage, 30,* 1255–1272.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). New York: Springer Verlag.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. J. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Computational Biology, 4,* e1000092. doi:10.1371/journal.pcbi.1000092.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Foldiak, P. (1990). Forming sparse representation by local Hebbian learning. *Biological Cybernetics, 64,* 165–170.

Friston, K. J. (2002a). Bayesian estimation of dynamical systems: An application to fMRI. *Neuroimage, 16,* 1325–1352.

Friston, K. J. (2002b). Beyond phrenology: What can neuroimaging tell us about distributed circuitry. *Annual Review of Neuroscience, 25,* 221–250.

Friston, K. J. (2005). Theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 360,* 815–836.

Friston, K. J. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4,* e1000211. doi 10.1371/journal.pcbi.e1000211.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19,* 1273–1302.

Friston, K. J., & Kiebel, S. (2009). Predictive coding under the free energy principle. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 364,* 1211–1221.

Galaburda, A., & Snides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190,* 597–610.

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *Neuroimage, 42,* 936–944.

Garrido, M. I., Kilner, J. M., Keibel, S., & Friston, K. J. (2009). Dynamic causal modelling of the response to frequency deviants. *Journal of Neurophysiology, 101,* 2620–2631.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology, 20,* 1–5.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *Neuroimage, 15,* 207–216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage, 22,* 755–766.

Hackett, T. A. (2007). Organization and correspondence of the auditory cortex of humans and nonhuman primates. In J. H. Kaas (Ed.), *Evolution of the nervous system* (pp. 109–119). Oxford, UK: Elsevier.

Hall, D., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex, 19,* 576–585.

Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS One, 5,* e9926.

Howard, M. A., Volkov, I. O., Granner, M. A., Damasio, H. M., Ollendieck, M. C., & Bakken, H. E. (1996). A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. *Journal of Neurosurgery, 84,* 129–132.

Jansen, B. H., & Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics, 73,* 357–366.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisons of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences, U.S.A., 24,* 11793–11799.

Kersten, D., Mammasian, P., & Yuille, A. (2004). Object perception and Bayesian inference. *Annual Review of Psychology, 55,* 271–304.

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., & Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Computational Biology, 5,* e1000464.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschls gyrus. *Cerebral Cortex, 13,* 765–772.

Maunsell, J. H. R., & Van Essen, D. C. (1983). The connections of the middle temporal visual area in the macaque monkey and their relationship to a hierarchy of cortical areas. *Journal of Neuroscience, 3,* 2563–2586.

Moran, R. J., Stephan, K. E., Seidenbecher, T., Pape, H. C., Dolan, R. J., & Friston, K. J. (2009). Dynamic causal models of steady state responses. *Neuroimage, 44,* 272–284.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage, 13,* 684–701.

Mumford, D. (1992). On the computational architecture of the neocortex: II. The role of cortico-cortical lops. *Biological Cybernetics, 66,* 241–251.

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated ripple noise. *Journal of the Acoustical Society of America, 100,* 3286–3294.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36,* 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience, 24,* 6810–6815.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of dynamic models. *PLoS Computational Biology, 6,* e1000709, doi:10.1371 /journal.pcbi.1000709.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage, 49,* 1641–1649.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. *Nature Neurocience, 2,* 79–87.

Reddy, C. G., Dahdaleh, N. S., Albert, G., Chen, F., Hansen, D., Nourski, K. V., et al. (2010). A method for placing Heschl gyrus depth electrodes. *Journal of Neurosurgery, 112,* 1301–1317.

Sirosh, J., & Mikkulainen, R. (1996). Self-organization and functional role of lateral connections and multisize receptive

fields in the primary visual cortex. *Neural Processing Letters, 3,* 39–48.

Stephan, K. E., Penny, W. D., Duanizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage, 46,* 1004–1017.

von Economo, C., & Koskinas, G. (1925). *Die cytoarchitectonik der hirnrinde des erwachsenen menschen.* Wien: Springer.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music- Brain responses to rhythmic incongruent. *Cortex, 45,* 80–92.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and height in the human brain. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 10038–10042.

Yost, W. A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 511–518.

Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated ripple noise. *Journal of the Acoustical Society of America, 99,* 1066–1078.