Research paper

Approaches to the cortical analysis of auditory objects

Timothy D. Griffiths [a,b,*], Sukhbinder Kumar [a], Jason D. Warren [b], Lauren Stewart [b], Klaas Enno Stephan [b], Karl J. Friston [b]

[a] *Auditory Group, Medical School, Framlington Place, University of Newcastle, Newcastle upon Tyne NE2 4 HH, United Kingdom*
[b] *Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, London WC1N 6BT, United Kingdom*

Abstract

We describe work that addresses the cortical basis for the analysis of auditory objects using 'generic' sounds that do not correspond to any particular events or sources (like vowels or voices) that have semantic association. The experiments involve the manipulation of synthetic sounds to produce systematic changes of stimulus features, such as spectral envelope.

Conventional analyses of normal functional imaging data demonstrate that the analysis of spectral envelope and perceived timbral change involves a network consisting of planum temporale (PT) bilaterally and the right superior temporal sulcus (STS). Further analysis of imaging data using dynamic causal modelling (DCM) and Bayesian model selection was carried out in the right hemisphere areas to determine the effective connectivity between these auditory areas. Specifically, the objective was to determine if the analysis of spectral envelope in the network is done in a serial fashion (that is from HG to PT to STS) or parallel fashion (that is PT and STS receives input from HG simultaneously). Two families of models, serial and parallel (16 in total) that represent different hypotheses about the connectivity between HG, PT and STS were selected. The models within a family differ with respect to the pathway that is modulated by the analysis of spectral envelope. After the models are identified, Bayesian model selection procedure is then used to select the 'optimal' model from the specified models. The data strongly support a particular serial model containing modulation of the HG to PT effective connectivity during spectral envelope variation.

Parallel work in neurological subjects addresses the effect of lesions to different parts of this network. We have recently studied in detail subjects with 'dystimbria': an alteration in the perceived quality of auditory objects distinct from pitch or loudness change. The subjects have lesions of the normal network described above with normal perception of pitch strength but abnormal perception of the analysis of spectral envelope change.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Auditory cortex; Functional imaging; Neurology; Lesion

1. Introduction

The concept of auditory object is controversial (Griffiths and Warren, 2004). The term can be applied to a sound source such as a voice, or an acoustic event generated by a source such as a vowel sound. In both cases there are features of the object that are independent of the detailed structure of the sound: we can recognise the same vowel, or voice, regardless of the pitch. In these examples the spectral envelope of the sound determines the particular vowel sound produced, and is one characteristic of the voice. Recognising the vowel or voice requires the 'abstraction' of spectral envelope. We are carrying out experiments that address how the auditory cortex might carry out such abstraction at a 'generic' level of processing (Griffiths et al., 2004b), before semantic analysis. Fig. 1 shows stimuli used in an imaging experiment (Warren et al., 2005) to assess the abstraction of changing spectral envelope regardless of the continuously changing fine spectral structure. In this experiment the key contrast compared the brain activation during the perception of sounds with changing spectral

* Corresponding author. Address: Auditory Group, Medical School, Framlington Place, University of Newcastle, Newcastle upon Tyne NE2 4 HH, United Kingdom.
 E-mail address: t.d.griffiths@ncl.ac.uk (T.D. Griffiths).

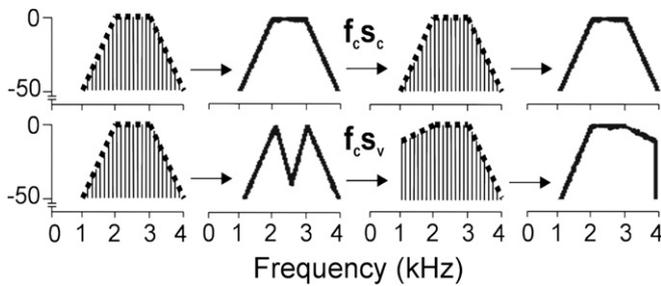Fig. 1. Stimuli for functional imaging of the brain basis for spectral-envelope analysis.

Fig. 2. Functional imaging of the brain basis for the analysis of spectral envelope.

envelope and brain activity during the perception of sounds with a fixed spectral envelope. In both conditions the individual sounds alternated between noise and harmonic stimuli, allowing an argument to be made that the difference between activation in the two conditions is due to the analysis of spectral envelope as opposed to the fine spectral structure.

Spectral analysis has been identified as an important aspect of auditory object characterisation in behavioural experiments that use the technique of multidimensional scaling (MDS) to place objects in a Euclidian space in which the distance between objects corresponds to the perceptual dissimilarity between them (Grey, 1977; McAdams and Cunible, 1992). The technique has typically been applied to the analysis of sources such as musical instruments and instrument hybrids. The studies generally agree that at least two dimensions are needed to characterise the timbre space, where these broadly correspond to the spectral and temporal envelopes, but the exact nature of the third dimension is not clear: early studies suggested a role for spectral flux (defined as the rate of change of spectral centroid) whilst later studies (Caclin et al., 2005) suggest that a role for fine spectral structure in object characterisation. The current approach to spectral envelope analysis is distinct to the approach using MDS based on 'real' objects. The approach might be described as the investigation of 'prototimbre': the systematic manipulation of higher-order properties of synthetic stimuli to investigate plausible bases for object characterisation.

2. Conventional mapping of cortical networks for the analysis of object properties

Brain activity during the analysis of spectral envelope was estimated by the fMRI BOLD response using a sparse imaging paradigm described in Warren et al. (2005). In Fig. 2 the critical contrast between changing and fixed spectral envelope, where the baseline fine spectral structure is continuously varying, reveals activity in a network of areas including planum temporale, PT (bilaterally), and superior temporal sulcus, STS (on the right). The STS activation occurs in a region that has been previously implicated in voice analysis (Belin et al., 2000), and can be interpreted in terms of a fundamental role in the abstraction of spectral
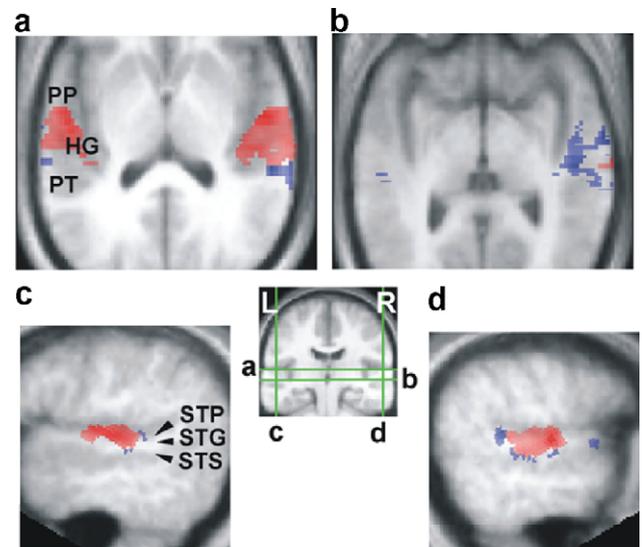
envelope relevant to the analysis of a number of sources including voices.

The experimental manipulation of spectral envelope in the above experiment can be regarded as manipulation of higher order properties of the stimuli that are perceived by the subjects, despite continuously changing fine spectro-temporal structure. From first principles, the activation network that is demonstrated might correspond to computation of higher order stimulus properties or to a neural correlate of the perceptual changes produced. We have demonstrated very similar networks of activation in experiments where distinct dimensions of timbre were manipulated (unpublished observation). The demonstration of a similar mapping for very different types of stimulus change is indirect evidence that activity in the network is a neural correlate of the changing percept. The argument is logically similar to one that we have developed related to the functional imaging of pitch perception, where manipulation of the stimulus in either the time domain (Patterson et al., 2002) or frequency domain (Penagos et al., 2004) produces similar changes in activity in lateral Heschl's Gyrus (HG), providing indirect evidence for the existence of a perceptual 'pitch centre' in lateral HG. With respect to the current network of activation involving bilateral PT and right lateralised STS it should be borne in mind that it is possible that different elements of the network have different relationships to stimulus computation or perception.

The lateralisation seen in the right temporal lobe in these experiments on normal subjects is congruent with data from the systematic study of patients with temporal lobectomy (reviewed in Stewart et al. (2006)) showing that right but not left temporal lobectomy affects the analysis of both spectral and temporal dimensions of timbre.

3. Beyond conventional mapping: dynamic causal modelling and Bayesian model selection

Conventional functional imaging analysis based on the general linear model is an exploratory approach that interrogates data from the whole brain to identify areas that show the specified relationship between activity and the stimulus manipulation or perceptual change. We have used dynamic causal modelling and Bayesian model selection approach to test different models that might account for the network of activity observed in the spectral envelope experiments. The aim of the approach is to address two fundamental biological questions. First we assess the general structure of the HG–PT–STS network for object processing. In particular, we address the critical question of whether analysis in PT and STS occurs in a serial (hierarchical) fashion, based on connections from HG to PT and from PT to STS, or whether the analysis is based on parallel processing due to connections from HG to both PT and STS. Second, we address how connection strengths between elements of this cortical network are modulated during the spectral envelope processing.

In order to test these hypotheses, two broad categories of models, serial and parallel were specified (Fig. 3). All the models specified were based on the (orthodox) assumption that there is a direct effect of the sound input on the activity within the primary auditory cortex within HG. In

the serial models, auditory inputs entering HG reach STS via PT and thus processing in STS depends on inputs from PT. In contrast, in the parallel models, HG connects to both PT and STS enabling parallel processing in PT and STS. In total, 16 models (9 serial, 7 parallel) were fitted to the data and compared using Bayesian model selection.

It should be noted that the DCM approach developed below might yield a 'best' model that is not a 'true' model if the set of models tested does not include the latter. It is important when setting up DCM analyses, therefore, to consider the possible models in a systematic and inclusive way. Even for a simple serial and parallel comparison for three areas as here, there are a large number of models when all the possible forward and back projections and the possible sites of modulatory effect are taken into account.

3.1. Dynamic causal modelling

The basic idea behind dynamic casual modelling can be summarised as follows. A cognitive or motor task in brain is accomplished by interaction between a number of nodes. This interaction is at the level of neural activity and therefore takes place at the millisecond time scale. DCM models such interaction. In an 'ideal' imaging technique that could accurately define nodes and the time course of neural activity at the millisecond level DCM could be carried out directly on the time series of neural activity from different nodes. Such an approach would actually be possible with neurophysiological data acquired simultaneously from different functional areas, though we are not aware of any studies that have used this approach. In humans, DCM has been applied to EEG and MEG data (Kiebel et al., 2006) with millisecond time resolution, but its main application to date has been to functional MRI data sets. For these applications, DCM incorporates a biophysical model for the conversion of regional neural activity into the 'sluggish' BOLD response that takes seconds to reach a peak in activation paradigms. Despite that fact that DCM is based on BOLD sampling rates that are more than three orders of magnitude slower than the temporal dynamics of the neural activity producing it, the incorporation of the biophysical model allows inference to be made from coarse resolution time series about events occurring at a much finer time scale. With respect to auditory functional MRI many experiments, including our experiment on spectral-envelope analysis above, are based on 'sparse' designs to avoid the effect of simultaneous scanner noise on the effects of interest. The current DCM analysis is the first to our knowledge that has been applied to such a data set. DCM allows inference about events occurring at the millisecond level based on BOLD time series acquired with sampling rates of less than 1 Hz. We will show in the analysis below that the additional loss of resolution due to sparse paradigms (where the sampling rate can approach 0.1 Hz) does not preclude the demonstration of plausible models of dynamic neural interaction using DCM.
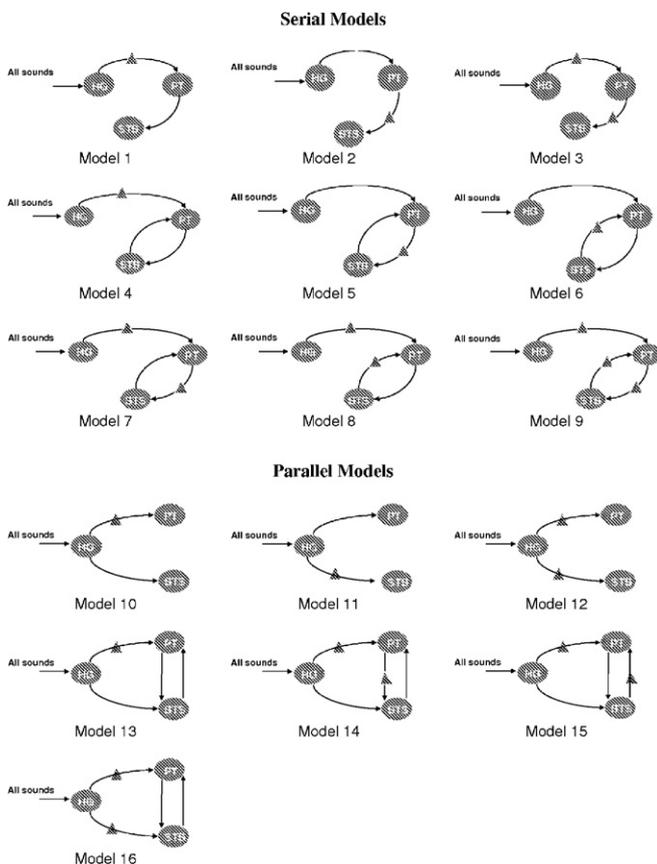

Fig. 3. Models for spectral envelope analysis in the right hemisphere.

From the system theory point of view, the brain can be treated as a nonlinear input–output dynamic system that can be excited by controlled stimuli and its response (hemodynamic response here) measured. The central idea behind dynamic causal modelling is to estimate and draw inference about the causal interaction between different regions of the brain by identifying a model for the system using input–output measurements.

In general, any model will comprise variables (that may or may not be measurable) and parameters that are estimated from the measurements. The model used in DCM has three types of variables: input variables that are the same as those used in conventional GLM analysis and encode the experimental manipulation; output variables that are the regional hemodynamic responses from each of the region considered in the model and state variables. State variables can often not be measured directly, and represent the neural activity and biophysical variables that transform neural activity into hemodynamic response.

In DCM, three different sets of parameters are used. The first set of parameters, known as intrinsic parameters, models the anatomical or hardwired connection strengths between the regions. These parameters represent the influence that one region has over the other in the absence of any external excitation of the system. The second set of parameters, known as modulatory parameters, models the change in intrinsic connection strength that is induced by the external experimental input. These parameters are therefore input specific and are also referred to as 'bilinear terms or parameters'. The third set of parameters models the direct influence of an external stimulus on a given region. The conventional GLM analysis is based on the assumption that any external stimulus has a direct influence on a region and therefore, it is the third set of parameters that form the primary focus of GLM analysis. DCM, therefore, can also be regarded as more general with GLM analysis being a specific situation where the interaction parameters (first and second set) are assumed be zero.

Once the model is specified, it has to be estimated from the measurements. There are, however, some natural constraints on the model. For example, the neural activity of a region can not diverge to infinity. The model parameters, therefore, need to be estimated such that these constraints are complied with. One framework for estimating the parameters with prior constraints is Bayesian statistics. In Bayesian statistics, a parameter is assumed as a random variable and is therefore completely characterized by its probability density (distribution) function. The prior constraints about the parameters are specified in terms of (prior) density function. Bayesian estimation procedure estimates the parameters by computing its posterior density function. Under the assumption that the density function is Gaussian, the density function can be characterized by two parameters namely maximum a posterior (MAP) estimate (which is equal to the mean value of the posterior density function under Gaussian assumption) and a posterior covariance. Once the probability density function of the parameters is known, meaningful inferences about the parameters can be drawn.

DCM has several advantages over other models of effective connectivity, e.g. structural equation modelling (SEM) (McIntosh and Gonzalez-Lima, 1994), multivariate autoregression (MAR) (Harrison et al., 2003) or Granger causality (Goebel et al., 2003) (see (Friston et al., 2003; Stephan, 2004) for further discussion). DCM takes into account temporal order (and autocorrelation of the fMRI time series). It further allows one to model the effects of experimentally controlled manipulations as either affecting regional activity directly (e.g. sensory inputs) or modulating the strengths of connections. Most importantly, however, DCM is currently the only model of effective connectivity that combines a neural population model with a biophysical hemodynamic forward model and is thus able to model how system dynamics at the (hidden) neuronal level translates into measured BOLD signals. These are the first data of which we are aware that apply the approach to sparse data sets from auditory experiments.

Mathematically, DCM is based on a bilinear model of neural population dynamics that is combined with a hemodynamic model (Buxton et al., 1998; Friston et al., 2000), describing the transformation of neural activity into predicted BOLD responses. The neural dynamics are modelled by the following bilinear differential equation

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Az + \sum_{j=1}^{m} u_j B^{(j)} z + Cu \qquad (1)$$

where z is the state vector (with one state variable per region), t is continuous time, and u_j is the jth input to the modelled system (i.e. some experimentally controlled manipulation). This state equation represents the strength of connections between the modelled regions (the A matrix), the modulation of these connections as a function of experimental manipulations (e.g. changes in task; the $B^{(1)} \ldots B^{(m)}$ matrices) and the strengths of direct inputs to the modelled system (e.g. sensory stimuli; the C matrix). These parameters correspond to the rate constants of the modelled neurophysiological processes. Combining the neural and hemodynamic model into a joint forward model, DCM uses a Bayesian estimation scheme to determine the posterior density of the parameters. Under Gaussian assumptions, this density can be characterised in terms of its maximum a posteriori (MAP) estimate and its posterior covariance.

Overall, the parameters of the neural and hemodynamic model are fitted such that the modelled BOLD signals are as similar as possible to the observed BOLD responses. This allows one to understand and make statistical inferences about regional BOLD responses in terms of the connectivity at the underlying neural level.

3.2. Selection of optimal model

A general problem that arises in any modelling exercise is to decide, given a measured data set, which of several competing models is the optimal. A number of criteria (for selecting the optimal model) have been proposed in the modelling the literature (Burnham and Anderson, 2004). From a Bayesian perspective, an optimal criterion is the model evidence, i.e. the probability $p(y|m)$ of obtaining the data y given a particular model m (Raftery, 1995). Critically, the model evidence not only takes into account the relative fit of competing models but also their relative complexity (i.e. the number of free parameters). This is important because there is a trade-off between the fit of a model and its generalizability, i.e. how well it explains different data sets generated from the same underlying process. As the number of free parameters is increased, model fit increases monotonically whereas beyond a certain point model generalizability decreases. The reason for this is "overfitting": an increasingly complex model will, at some point, start to fit noise that is specific to one data set and thus become less generalizable across multiple realizations of the same underlying generative process.

As the model evidence cannot always be derived analytically, two commonly used approximations to it are Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Penny et al., 2004). Application of these approximations for model selection, however, do not necessarily give identical results because BIC favours simpler model whereas AIC is biased towards more complex models. A general convention is that if two models (say m_1 and m_2) are to be compared, then a decision is made only when AIC and BIC concur. The relative evidence of one model as compared to another is determined using the so-called 'Bayes factor':

$$BF_{12} = \frac{p(y|m_1)}{p(y|m_2)} \tag{2}$$

where BF_{12} is the Bayes factor of model 1 with respect to 2. Following the selection of a best model for each individual subject, the optimal model for a group of subjects can be determined by the group Bayes factor (GBF), which is equal to the geometric mean of the Bayes factors for each individual subject.

3.3. The optimal model for spectral analysis

Fig. 4 shows the evidence for the models, determined separately using AIC and BIC, in one subject. Model 1 is the optimal model to explain the data. The parameters for this model specify a serial model with connectivity (HG → PT → STS) and modulation of connection from HG to PT during the analysis of spectral envelope. In addition to the individual inference, Table 1 shows the Bayes factor (minimum of the two values computed using AIC and BIC) for model 1 with respect to the other 15 models. All the values are greater than 150, except for models 8 and
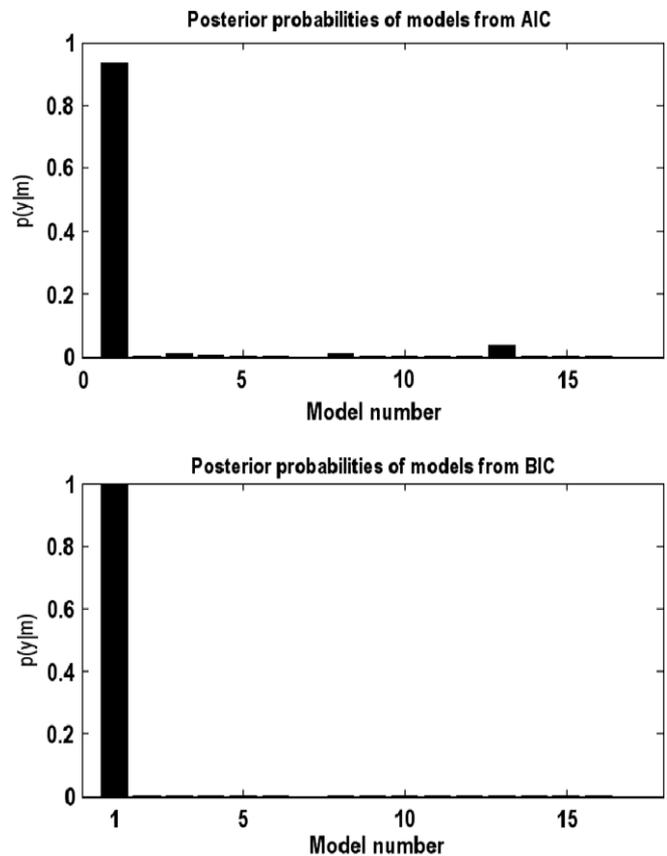

Fig. 4. Evidence for the models in a single subject: Plots of probabilities $p(y|m)$ for individual subjects for the 16 models included in dynamic causal modelling. The probabilities have been normalised so that they sum to one. These represent the probability of the model, given the data, assuming each model is, a priori, equally likely.

Table 1
Bayes factor of model 1 vs. all the remaining models

Model 1 vs.	Bayes factor
2	6.27×10^7
3	1.06×10^2
4	1.50×10^2
5	1.14×10^{14}
6	1.01×10^4
7	3.68×10^{23}
8	0.85×10^2
9	8.64×10^9
10	6.85×10^5
11	1.64×10^7
12	1.75×10^7
13	0.26×10^2
14	4.21×10^5
15	9.13×10^2
16	2.84×10^2

13 for which Bayes factors are 85 and 26, respectively. This corresponds to 'strong' evidence for serial model with respect to models 8 and 14 and 'very strong' evidence with respect to all other models (Raftery, 1995).

Estimates were derived for the intrinsic and modulatory connection strengths (Table 2) of the optimal model and the probabilities that the parameter estimates are greater

Table 2
Intrinsic and modulator connection strengths for model 1 (optimal serial model)

Connection strength	HG–PT	PT–STS
Intrinsic	0.492(1.000)	0.525 (1.00)
Modulatory	0.619 (0.999)	–

than zero. The probabilities that the magnitudes of intrinsic connection strengths are greater than zero are significant Furthermore, the probability that the modulation of the strength of the connection from HG to PT is greater than is also significant.

The intrinsic modulations in the model correspond to effective connectivity: the direct modulatory effect of the activity in one area on that in another mediated by an anatomical connection. In anatomical terms such connection could be direct, or occur via a relay but there must be a mechanism for connection. Data about anatomical connections between human auditory areas are lacking: there are data showing connection for HG to PT (Tardif and Clarke, 2001) but we are not aware of any data on the connection from PT to STS predicted by the model. The basis for the modulatory connection deserves comment: this does not predict a direct anatomical effect on a connection, but rather a change in the connection strength between HG and PT during envelope analysis. The model predicts a component of the effect of activity in HG on that in PT that occurs specifically during spectral envelope analysis. This prediction could be tested with other methods such as neurophysiological recording from HG and PT during manipulation of the spectral envelope of stimuli.

The model has certain limitations and should not at this stage be regarded as a general synthesis of all aspects of object analysis in all subjects. Broadly, the existence of a serial model is in accord with the concept of a single pathway for auditory object analysis, and supports the concept of PT as a critical 'computational hub' (Griffiths and Warren, 2002) at the interface between the abstraction of auditory object properties and further analysis in distinct higher centres for object analysis that also carry out semantic level processing. Further work is required to asses the extent to which the model generalises to larger populations of subjects. It should also be emphasised that the approach taken here addresses the simplest level of perceptual analysis when the subject is required to attend to the sounds but does not carry out any relevant task or semantic level analysis. It will be of considerable future interest to examine the effects of task and semantic analysis on intrinsic connectivity patterns. A number of questions arise such as these levels of processing are associated with modulation at later stages of the system (the connection between PT and STS) or the existence of additional back connections that are predicted by models of visual cortical processing.

4. A cognitive neuropsychological approach

The approach developed here has the potential to allow a fuller understanding of human cortical disorders. First, disorders manifest at the level of ethological stimuli such as speech, environmental sounds and music can be considered in terms of more generic auditory processing to assess 'bottom-up' contributions to the deficit. The approach is developed in detail for music in Stewart et al. (2006) Second, an approach based on the exact specification of normal models will allow greater understanding of the effects of lesions in terms of their effects on connectivity between nodes in networks as well as the effects on the nodes themselves. Network models of auditory analysis incorporating connectivity go back to models such as the 1885 Lichteim model for speech (Lichtheim, 1885), based on an analysis
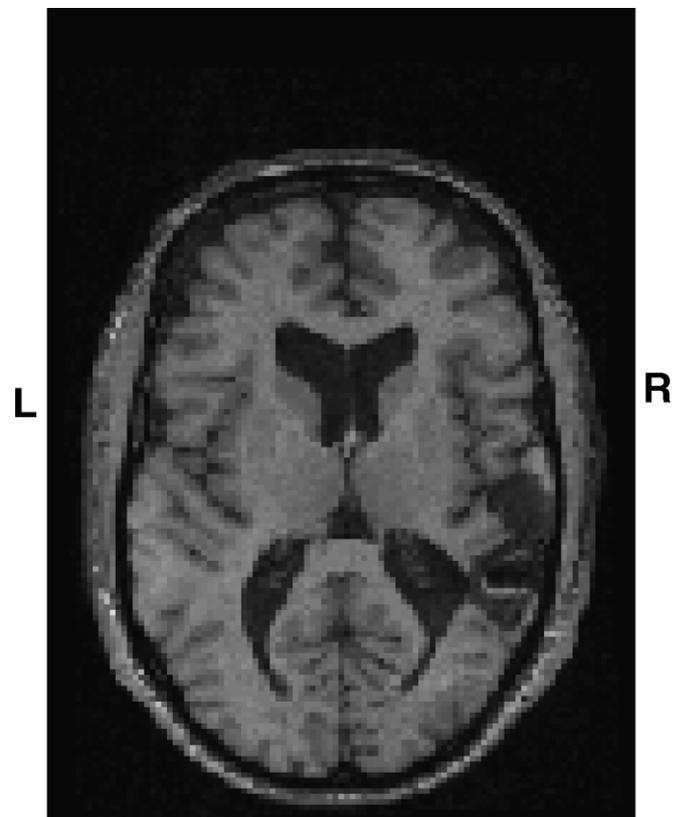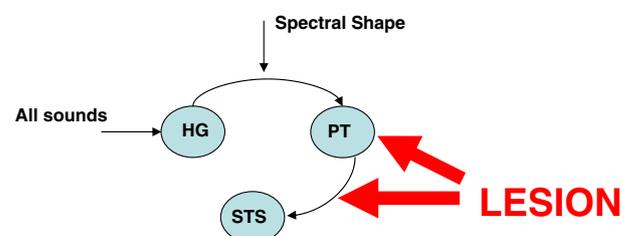

Fig. 5. Lesion in patient with dystimbria.

Fig. 6. Effect of lesion on the normal network for spectral envelope analysis.

Table 3
Psychophysical assessment of patient with 'dystimbria': normal perception of complex pitch and modulation and abnormal perception of timbral changes

Task	Control mean (SD)	Patient	Z score
Complex pitch			
Iterated rippled noise (gain)*	−1.05 (0.176)	−1.29	−1.41
Modulation			
2 Hz FM*	0.0136 (0.183)	0.154	0.767
120 Hz FM	0.0161 (0.00758)	0.0184	0.303
2 Hz AM*	−1.02 (0.264)	−1.10	−0.300
120 Hz AM*	−1.85 (0.249)	−1.71	0.553
Spectral timbre			
M timbre gain	7.6 dB (.649)	14.2	10.2
M timbre attenuation	7.5 dB (1.10)	13.3	5.3

The data for complex pitch and modulation are based on the exact methods specified in Griffiths et al. (2001) using two-interval-two-alternative forced choice psychophysics and method of constant stimuli to yield 75% thresholds that are compared with those of 20 age matched controls to yield a z score. * Show tests for which log transformation of the data is required to allow parametric statistics.
The data for spectral timbre are based on the detection of changes to the shape of an 'M' shaped spectral envelope applied to a harmonic series in positive Shroeder phase with fundamental frequency of 100 Hz. The stimulus was roved over 20 dB and subjects had to detect increased gain (more dip at central point) or attenuation (less dip). The reference stimulus had attenuations in the envelope of 50, 0, 15, 0, and 50 dB at 0.8, 1.6, 2.4, 3.2, and 4 kHz, respectively. An AXB design was used where subjects had to detect changes in the first or second stimulus compared to the middle reference. Threshold was derived from three-down-one-up adaptive tracking and z scores based on comparison with control data.

of the effect of stroke on the speech network: here abnormal function allows inference about normal function and the approach is still valid today (Peretz and Coltheart, 2003). We now have the ability to define the normal networks for auditory analysis with much more precision to allow prediction of the effect of lesions, so that knowledge of normal function can now allow inference about abnormal function.

Fig. 5 shows the structural damage to the brain of a patient who presented at the age of 42 with an unusual auditory deficit after a right-hemisphere stroke. Damage occurred to the lateral part of HG on the right, the right planum temporale and the posterior superior temporal gyrus and the adjacent posterior superior temporal sulcus. He presented with unusual symptoms that might be characterised as 'dystimbria': musical instruments appeared distorted to him, speech appeared distorted to him (human voices sounded 'unreal' as if they were being played through poor quality speakers) and he sometimes misidentified environmental sounds such as birdsong.

A problem with spectral envelope analysis in this patient would be predicted on the basis of model 1 in the analysis above. Fig. 6 shows the effect of the lesion on the spectral envelope system that would be predicted entirely on the basis of anatomy. The lesion in his case would affect both the PT node and the connection between PT and STS.

Psychophysical assessment of the patient (Table 3) confirmed normal perception of complex pitch as assessed by the detection of the pitch of individual sounds. Testing the detection of pitch in regular-interval noise with continuously variable gain to change regularity and pitch strength (using two-interval-two-alternative forced choice psychophysics and the method of constant stimuli) as described in Griffiths et al. (2001) yielded normal thresholds: this would be predicted on the basis of the intact pitch centre in lateral HG in the left hemisphere. He also demonstrated

normal performance for the detection of amplitude and frequency modulation of a narrowband carrier with rates of 2 Hz and 120 Hz. These thresholds are expressed in Table 3 as z-scores compared to a control age-matched population.

In contrast his thresholds for the detection of changes in spectral shape were double that of normal controls (Table 3). He had similar deficits in the detection of the frequency shift of the spectral envelope of a harmonic carrier and in the detection of changes in attack rate.

5. Conclusion

We have developed an approach to the analysis of auditory object properties based on the precise definition of the underlying cortical systems in normal subjects. The approach suggests serial analysis of the property of spectral envelope in HG then PT and STS and predicts deficits in timbral analysis that we have identified in patients with right-hemisphere lesions.

References

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Method Res. 33, 261–304.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magnet. Reson. Med. 39, 855–864.

Caclin, A., McAdams, S., Smith, B.K., Winsberg, S., 2005. Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J. Acoust. Soc. Am. 118, 471–482.

Friston, K.J., Mecheli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other haemodynamic. NeuroImage 12, 466–477.

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

Goebel, R., Roebroeck, A., Kim, D.S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modelling and Granger causality mapping. Magn. Reson. Imaging 21, 1251–1261.

Grey, J.M., 1977. Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am. 61, 1270–1277.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.

Griffiths, T.D., Warren, J.D., 2004. What is an auditory object?. Nature Rev. Neurosci. 5, 887–892.

Griffiths, T.D., Dean, J.L., Woods, W., Rees, A., Green, G.G.R., 2001. The Newcastle Auditory Battery (NAB) – a temporal and spatial test battery for use on adult naive subjects. Hear. Res. 154, 165–169.

Griffiths, T.D., Warren, J.D., Scott, S.K., Nelken, I., King, A.J., 2004b. Cortical processing of complex sound: a way forward?. Trends Neurosci. 27, 181–185.

Harrison, L.M., Penny, W., Friston, K.J., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage 19, 1477–1491.

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.

Lichtheim, L., 1885. On aphasia. Brain 7, 433–484.

McAdams, S., Cunible, J.C., 1992. Perception of timbral analogies. Phil. Trans. Royal Soc. – Series B: Biol. Sci. 336, 383–389.

McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modelling and its application to network analysis in function brain mapping. Hum. Brain Mapp. 2, 2–22.

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. J. Neurosci. 24, 6810–6815.

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

Peretz, I., Coltheart, M., 2003. Modularity of music processing. Nature Neurosci. 6, 688–691.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–196.

Stephan, K.E., 2004. On the role of general system theory for functional neuroimaging. J. Anat. 205, 443–470.

Stewart, L., Kriegtein, K., Warren, J.D., Griffiths, T.D., 2006. Music and the brain: disorders of musical listening. Brain. 129, 2533–2553.

Tardif, E., Clarke, S., 2001. Intrinsic connectivity in human auditory areas: tracing study with DiI. Eur. J. Neurosci. 13, 1045–1050.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. NeuroImage 24, 1052–1057.