Encoding of the temporal regularity of sound in the human brainstem

Timothy D. Griffiths[1,2,3], Stefan Uppenkamp[1], Ingrid Johnsrude[4], Oliver Josephs[2] and Roy D. Patterson[1]

[1] *Centre for the Neural Basis of Hearing, Physiology Department, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK*

[2] *Wellcome Department of Cognitive Neurology, Institute of Neurology, 12 Queen Square, London, WC1N 3BG, UK*

[3] *Auditory Group, Department of Physiological Sciences, Newcastle University Medical School, Newcastle upon Tyne, NE2 4HH, UK*

[4] *MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, UK*

The first two authors contributed equally to this work

Correspondence should be addressed to T.D.G. (t.d.griffiths@ncl.ac.uk)

We measured the neural activity associated with the temporal structure of sound in the human auditory pathway from cochlear nucleus to cortex. The temporal structure includes regularities at the millisecond level and pitch sequences at the hundreds-of-milliseconds level. Functional magnetic resonance imaging (fMRI) of the whole brain with cardiac triggering allowed simultaneous observation of activity in the brainstem, thalamus and cerebrum. This work shows that the process of recoding temporal patterns into a more stable form begins as early as the cochlear nucleus and continues up to auditory cortex.

Temporal processing in the auditory system is unique because resolution is at least an order of magnitude greater than in any other system. In the cochlea, after the basilar membrane performs an initial frequency analysis, inner hair cells and primary auditory nerve fibers along the edge of the basilar membrane convert the motion into action potentials in the auditory nerve fibers[1]. Such firing is 'phase locked' to membrane motion up to frequencies as high as 5000 Hz (ref. 2); that is, the fibers encode membrane peak times with sub-millisecond accuracy. Such phase locking is not observed in the higher nuclei of the ascending auditory pathway[3], which immediately raises the question of how temporal patterns at the millisecond level are encoded at higher points in the system. A representation of these temporal patterns in the higher points in the pathway is likely to be necessary for a number of perceptual phenomena including pitch perception[4,5] and voice identification[6].

Several models for the representation of temporal patterns in the auditory system have been proposed. Many use autocorrelation or coincidence detection to encode the regularity in the auditory signal[7,8]. In one model of human perception, a form of autocorrelation is carried out within each frequency band to produce an auditory pattern or 'image' that corresponds to a given percept[8]. Animal neurophysiology suggests that maps of periodicity probably exist in the midbrain[9], orthogonal to the well-known frequency map. However, there are important differences between the structure of the human ascending auditory pathway and that of other mammals[10]. Here we use functional MRI to investigate the basis for temporal processing in the human ascending pathway. We seek local brain activity that increases with the degree of regularity in the stimulus (the degree to which a repeating pattern over time at the millisecond level is present in the stimulus). Such a demonstration indicates the presence of a mechanism that converts temporal regularity into a local neuronal activity level; it does not specify the mechanism by which that conversion occurs.

We previously delimited the site at which temporal regularity is converted into an increase in neural activity by manipulating noise to enhance the probability of one particular time interval occurring in the neural firing pattern and using PET scanning to search for centers where activity was responsive to changes in the temporal regularity[11]. We use similar stimuli in the current study. Figure 1 illustrates how the temporal regularity of neural firing can be systematically varied without changing the distribution of energy across frequency. The simulated time-interval histogram or auditory image (corresponding to firing times in the frequency channels of the auditory nerve) is stabilized whenever the sound is periodic[4,8]. Increasing temporal regularity in the noise stimulus leads to the emergence of peaks in the auditory images. In contrast, no pattern emerges in the auditory spectra showing the overall activity levels in the same auditory nerve fibers. These stimuli, therefore, enable the investigation of the effect of auditory temporal structure on brain activity without a confounding change in the auditory spectrum. We have used these stimuli with an fMRI protocol designed to optimize visualization of the brainstem during auditory processing. This has allowed us to search simultaneously in the brainstem, thalamus and cortex for activity related to the processing of temporal regularity.

RESULTS

BOLD (blood oxygen level-dependent) responses to acoustic stimuli were measured in the whole brain using cardiac triggering to overcome the effect of pulsation of the basilar artery on the brainstem signal[12]. The sounds were delivered using a high-fidelity electrostatic system (Palmer, A. R. *et al*. *Neuroimage* 7, S359, 1998) and sparse imaging was used to remove the confounding effect of scanner noise on the brain activity[13]. Sparse imaging is a technique that exploits the sluggishness of the BOLD response (~10 s in auditory cortex[13]) to separate the effects of scanner noise from the response to the experimental stimuli.

Table 1. Coordinates of structures in the ascending pathway.

Structure	CN-l			CN-r			IC-l		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
Mean ± s.d. structural coordinates	–8 ±1.3	–41 ± 1.1	–47 ± 1.8	9 ± 1.1	–41 ± 1.1	–48 ± 1.6	–5 ± 0.5	–35 ± 0.9	–10 ± 1.3
Functional coordinates (group, sound-minus-silence contrast)	–12	–40	–46	8	–34	–48	–6	–34	–12
Group contrast significance (sound minus silence)		< 0.005			< 0.005			< 0.001	
Group contrast significance (regular-interval-sound minus noise)		< 0.05			< 0.05			< 0.005	
Group contrast significance (parametric analysis for 3 levels of regularity)		< 0.05			< 0.01			< 0.01	
Group contrast significance (lively-pitch minus fixed-pitch)		ns			ns			ns	

Coordinates as determined from individual structural data on eight subjects and from the group sound-minus-silence comparison, with significance levels for the contrasts between sound conditions. The coordinates are in mm in Talairach space. Significance levels are given after correction for multiple comparisons within the volume of interest defined (a 5-mm sphere centered on the functional coordinates).

Long stimulus presentations are used and single brain volumes are acquired after each presentation to measure the brain activity corresponding to stimulus alone before the scanning sequence.

Eight subjects underwent 288 scans each following the presentation of silence, Gaussian noise or sound with fixed temporal regularity. The degree of temporal regularity in the sound stimuli was systematically varied between scans. For the sound-minus-rest functional comparison, a volume-of-interest method was used to identify activations in the region of the cochlear nuclei (CN), inferior colliculi (IC) and medial geniculate bodies (MGB). These regions were defined by systematic inspection of the structural MRI for each subject using predefined anatomical criteria described in the Methods. The mean coordinates of the center of each structure in each subject were then used to define the center of a volume of interest for that structure for the group analysis. There was good correspondence between the structures of the ascending pathway defined by anatomical criteria and the local maxima for the sound-minus-rest contrast (Table 1). However, functional data (for the group sound-minus-rest contrast) were used in preference to the anatomical criteria to define the volumes of interest for the group analyses of the effect of temporal structure, because of a degree of uncertainty about the position of certain structures, especially MGB.

Comparing the activity during sound stimulation and silence revealed a pattern of strong activation largely confined to the auditory system (Fig. 2). Activation in CN, IC, MGB and the primary auditory cortices was highly significant bilaterally (Table 1, Fig. 2). Significant increases in BOLD response with the introduction of temporal regularity were demonstrated in

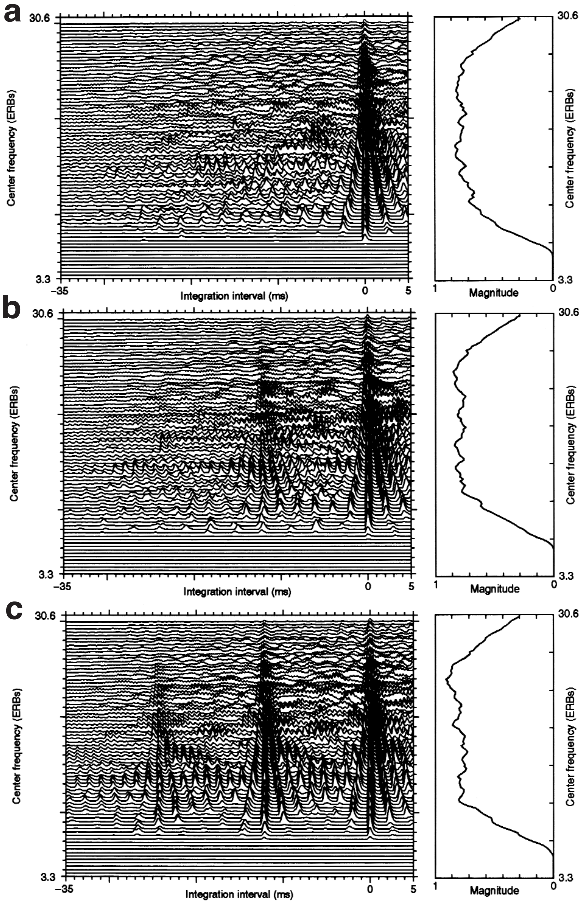

Fig. 1. Experimental protocol. Examples of the internal representations of regular-interval noise from a time-domain model (left) and a spectral model (right). Top, response to a noise stimulus (no temporal regularity and no pitch); middle, response to an intermediate degree of temporal regularity and a weak pitch; bottom, response when there is a high degree of temporal regularity and a strong pitch. Right panel in each row shows sum of the neural activity over time in each channel (a spectral model); no clear pattern of energy peaks emerges with the increase in the temporal regularity of the sound. Left panel in each row shows the 'auditory image' from a time domain model in which the temporal regularity is measured by a form of autocorrelation. A vertical structure centered at the delay (12 ms) emerges as the temporal regularity increases. ERB, equivalent rectangular bandwidth (absolute frequency scale based on the ear's frequency analysis).

Table 1, cont'd

	IC-r			MGB-l			MGB-r		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
Mean ± s.d. structural coordinates	5 ± 0.7	–35 ± 0.9	–10 ± 1.3	–16 ± 1.4	–26 ± 1.8	–8 ± 1.9	16 ± 1.1	–26 ± 2.3	–6 ± 2.2
Functional coordinates (group, sound-minus-silence contrast)	6	–36	–10	–16	–28	–10	10	–32	–8
Group contrast significance (sound minus silence)		< 0.001			< 0.001			< 0.005	
Group contrast significance (regular-interval-sound minus noise)		< 0.005			< 0.05			< 0.005	
Group contrast significance (parametric analysis for 3 levels of regularity)		< 0.005			ns			< 0.005	
Group contrast significance (lively-pitch minus fixed-pitch)	ns				ns			ns	

both CNs, both ICs and both MGBs (Table 1). The effect of the introduction of temporal regularity was assessed by the contrast between the BOLD response to all the stimuli with temporal regularity and the response to the noise control stimulus. Moreover, the study included three levels of temporal regularity, which enabled a parametric analysis. This demonstrated a positive correlation between activation and regularity in CN and IC bilaterally (Table 1). In these areas, brain activity increases when the temporal regularity is increased, despite the matching of the auditory spectra of the stimuli.

The stimuli containing temporal regularity were sequences of notes played at the rate of four per second. In one condition, the pitch was fixed, and in others, the pitch was varied to produce 'lively' note sequences. Comparison between fixed and lively pitch sequences did not produce any differences in the BOLD responses in auditory brainstem nuclei (Table 1). In the cortex, in contrast, a distinct pattern of activation was observed (Fig. 3). Bilateral activation was demonstrated in lateral Heschl's gyrus in an area of secondary auditory cortex, in addition to bilateral activation posterior to Heschl's gyrus in the planum temporale.

We carried out a further analysis to test explicitly the hypothesis that IC is more sensitive to changes in temporal regularity than CN. For each subject, the effect of regularity on the mean regional activity in the CN and IC was estimated using the local eigenvariate (averaged across the two sides) for the contrast between regular-interval stimuli and noise. The effect of regularity on activity for the CN and IC was then compared using the Wilcoxon matched-pairs signed-rank test to compare the group eigenvariates for the two centers. This analysis amounts to seeking an interaction between BOLD response and regularity at the two centers, and demonstrated a significant increase in sensitivity of the IC to changes in temporal regularity ($p < 0.05$, two-tailed test).

DISCUSSION

We demonstrated that temporal regularity in a sound affects the BOLD response in the first processing stage of the auditory pathway, the CN. As the BOLD response due to an ensemble of neurons is likely to reflect the average neuronal firing rate[14], this implies that the temporal regularity in a sound affects the average neuronal firing rate in the CN. Such an increase in firing rate might come about in several ways. First, specific neurons that are sensitive to the particular range of periodicities of the stimulus could show an increase in firing rate. Previous animal work shows that onset-chopper cells in the ventral CN respond preferentially to stimuli with a particular periodicity[15]. Alternatively, an increase in the synchronization of CN neuron firing could produce an increase in mean firing rate. (See ref. 16 for modeled networks showing tight coupling between firing rate and synchronization.) This increase in synchronization could occur in a number of different cell types in the ventral CN.

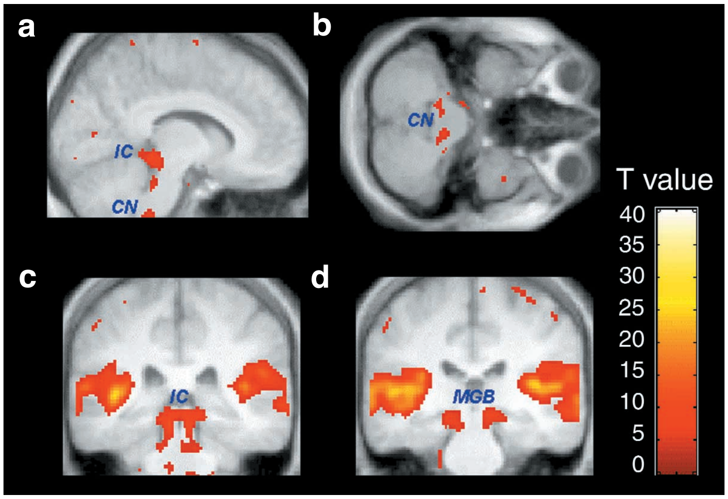

Fig. 2. Activation of structures in the ascending auditory pathway with sound stimuli using cardiac triggering and sparse imaging (contrast between all sound stimuli and silence shown in relation to average structural MRI). Sagittal section at $x = 10$ mm showing activation in the right cochlear nucleus and inferior colliculus (**a**), axial section at $z = -46$ mm showing bilateral activation of cochlear nuclei (**b**), coronal section at $y = -34$ mm showing activation of inferior colliculi and superior temporal cortex (**c**), coronal section at $y = -28$ mm showing activation of medial geniculate bodies (**d**). Threshold for contrast $p < 0.001$ (uncorrected). Color scale gives Student's t statistic for the comparison between the BOLD values in the sound conditions and rest.

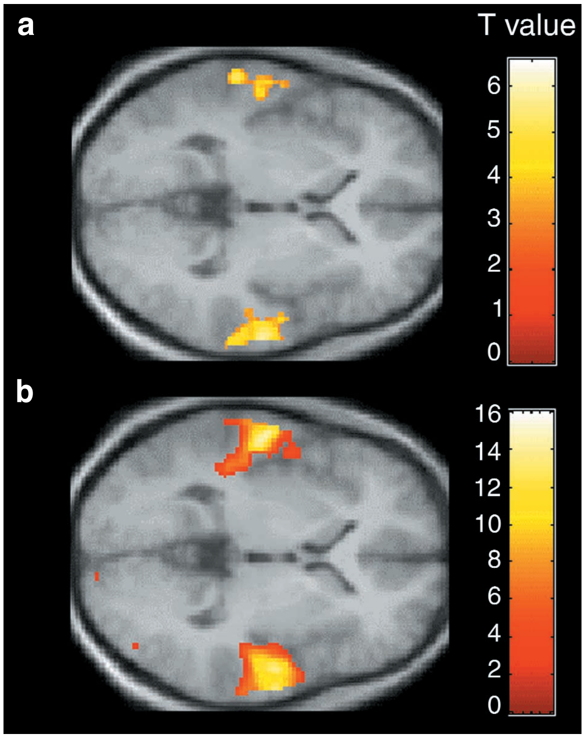

Fig 3. Activation during the processing of the pitch changes between sounds and the temporal structure within sounds. (**a**) Contrast between the 'lively' pitch sequences and a fixed pitch sequence. Axial section at $z = 0$ mm. Significant activation is shown in secondary auditory cortex in lateral Heschl's gyrus and cortex posterior to Heschl's gyrus in the plana temporale. No significant activation differences were demonstrated in the brainstem or thalamus. (**b**) For comparison, the contrast between regular interval sound and noise at the same axial level. This shows activation extending along the whole of Heschl's gyrus including primary auditory cortex in the medial part. Threshold for contrasts, $p < 0.001$ (uncorrected).

The relationship between the BOLD response and temporal regularity is even more significant in the IC ($p < 0.005$ in IC compared to $p < 0.05$ in CN), and the IC has been shown to be significantly more sensitive to changes in regularity than the CN. If the activity in the CN represents conversion of temporal regularity into an activity code (rather than some aspect of synchronization), this would suggest that the conversion starts in CN but is not completed there. For example, it is possible that time intervals between neural impulses are actively measured in the CN but that the time-interval histograms where the effect accumulates are in the IC, as suggested for the delay maps of Langner[9]. The early measurement of time intervals in the CN would minimize the loss of timing information that occurs between the CN and IC. Alternately, the activity in the CN may correspond to a representation of the envelope peaks (a passive accentuation of the synchronization rather than active measurement). This representation of envelope peaks in the CN could provide a basis for time-interval measurement in the IC, as proposed in the auditory image model[8]. In either case, this is the first evidence that the kind of two-stage physiological mechanism demonstrated in animals[9] and proposed in auditory models of perception[8] also exists in humans.

In contrast to the striking effect of temporal structure at the millisecond level, no change in brainstem activity is observed when the pitch changes at rates common in music and speech. Moreover, altering these longer-term temporal patterns does not produce any activity differences in primary auditory cortex. Such differences only emerge in secondary auditory cortex. Previous animal studies have documented the importance of secondary auditory areas in the processing of complex sounds characterized by relatively slow changes in temporal features over time[17]. Melody processing in humans, in particular, is associated with right-lateralized activity in secondary auditory cortex and the planum temporale[18,19].

These findings are consistent with hierarchical processing of the temporal structure of sound[11].

METHODS

Stimuli and protocol. Regular interval sounds, in which the temporal regularity is varied systematically[11], were used. Perceptually, these sounds produce a well-defined pitch appropriate to the delay, and the strength of the pitch increases with the degree of temporal regularity in the neural activity. Theoretically, they are important because they produce pitch without having the appropriate set of auditory spectral peaks[4,20]. The sounds were delivered using a high-fidelity electrostatic sound system to preserve the temporal and spectral characteristics.

Individual sounds, 200-ms in duration, were produced by a delay-and-add algorithm applied to broadband noise that was then filtered using fourth-order Butterworth filters between 500 Hz and 4 kHz. Each delay-and-add cycle is called an iteration; the delay-and-add process increases the temporal regularity of the sound and, thus, the pitch strength. Low pitch values and high-pass filtering of the stimulus minimize any resolvable spectral change due to the delay-and-add process. Zero, 2 and 16 iterations were used, and the individual pitch values were always between 50 and 110 Hz. Sequences of sounds with different pitches were constructed from the individual sounds. In the sequences, the pitch values of the individual sounds were varied in three different ways: there was a fixed-pitch condition, a randomized-pitch condition and a novel-melody condition. In any given sequence, the degree of temporal regularity in the individual sounds and the pitch strength were fixed. The randomized pitch and melody conditions are referred to as 'lively' in reference to the pitch excursions between successive notes.

Before acquisition of each brain image volume, subjects listened to silence or to one of the sound sequences. This allowed analysis of the effect on brain activity of varying the temporal regularity in the individual sounds and the effect of varying the pitch patterns formed by those sounds. The subjects were instructed to listen to any pattern that they could detect within the sounds.

Examples of stimuli are available on the *Nature Neuroscience* web site (http://neurosci.nature.com/web_specials).

fMRI acquisition. Sparse imaging[13] was used to enable examination of activation due to the stimuli, in the absence of the effect of the noise generated by the scanner. This is achieved through the infrequent acquisition of brain images just after lengthy periods of stimulus presentation, during which there is no image acquisition, and, thus, no scanner noise. Identical sound stimuli were presented to both ears at 75 dB sound pressure level (SPL). BOLD contrast image volumes were acquired at 2.0 T (Siemens, VISION, Erlangen) with gradient-echo-planar imaging (TR/TE, 12000 ms/35 ms). Subjects underwent three runs; each included 16 silent control scans and 80 scans following presentation of sound sequences.

Cardiac triggering[12] was used to overcome the degradation of brainstem image quality that can result from the effect of pulsation of the basilar artery. This is achieved by always acquiring a given image slice at the same point in the cardiac cycle. Acquisition of each brain volume was triggered by the first R wave of the recorded electrocardiogram occurring after 8 s of either silence or a sound sequence. Ascending axial acquisition was used (48 slices) to maximize the value of the cardiac triggering for the brainstem structures while allowing acquisition of whole brain images.

fMRI analysis. Analysis of BOLD image time series was carried out using SPM99 (http://www.fil.ion.ucl.ac.uk/spm) after realignment, normalization and smoothing with a Gaussian filter (full width at half maximum, 5 mm). A fixed effects analysis was carried out using the general linear model. For the volume-of-interest analysis, spheres of 5-mm radius were defined for the cochlear nuclei, inferior colliculi and medical geniculate bodies according to the criteria below. The analyses for the main effects of temporal structure and sequence 'liveliness' sought significant activation at the cluster level within these volumes.

Volumes of interest. Criteria for identification of structures in the ascending auditory pathway using a standard human atlas (for example, ref. 21) are available on the *Nature Neuroscience* web site (http://neurosci.nature.com/web_specials). For the CN, IC and MGB, the criteria were used to define the center of a spherical volume of 5-mm radius containing the structure in question.

RECEIVED 2 OCTOBER 2000; ACCEPTED 12 APRIL 2001

1. Nobili, R., Mammano, F. & Ashmore, J. How well do we understand the cochlea? *Trends Neurosci*. 21, 159–167 (1998).
2. Rose, J. E., Brugge, J. F., Anderson, D. J. & Hind, J. E. Phase-locked response to low-frequency tones in single auditory nerve fibres of the squirrel monkey. *J. Neurophysiol*. 30, 769–793 (1967).
3. Rouilly, E., deRibaupierre, Y. & deRibaupierre, F. Phase-locked responses to low frequency tones in the medial geniculate body. *Hear. Res*. 1, 213–226 (1979).
4. Yost, W. A., Patterson, R. D. & Sheft, S. A time domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am*. 99, 1066–1078 (1996).
5. Patterson, R. D., Yost, W. A., Handel, S. & Datta, A. J. The perceptual tone/noise ratio of merged iterated rippled noises. *J. Acoust. Soc. Am*. 107, 1578–1588 (2000).
6. Patterson, R. D., Anderson, T. & Allerhand, M. in *Proceedings of the International Conference on Spoken Language Processing 1994*, 1395–1398 (Yokohama, Japan, 1994).
7. Langner, G. Periodicity encoding in the auditory system. *Hear. Res*. 60, 115–142 (1992).
8. Patterson, R. D., Allerhand, M. H. & Giguère, C. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am*. 98, 1890–1894 (1995).
9. Langner, G. & Schreiner, C. E. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol*. 60, 1799–1822 (1988).
10. Moore, J. K. T. in *Neurotology* (eds. Jackier, R. K. & Brackmann, D. E.) 1–17 (Mosby, St. Louis, 1994).
11. Griffiths, T. D., Buechel, C., Frackowiak, R. S. J. & Patterson, R. D. Analysis of temporal structure in sound by the human brain. *Nat. Neurosci*. 1, 421–427 (1998).
12. Guimares, A. R. *et al*. Imaging subcortical activity in humans. *Hum. Brain. Mapp*. 6, 33–41 (1998).
13. Hall, D. A. *et al*. "Sparse" temporal sampling in auditory fMRI. *Hum. Brain. Mapp*. 7, 213–223 (1999).
14. Rees, G., Friston, K. & Koch, C. A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci*. 3, 716–723 (2000).
15. Wiegrebe, L. & Winter, I. M. Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. *J. Neurophysiol*. 85, 1206–1219 (2001).
16. Chawla, D., Lumer, E. D. & Friston, K. J. The relationship between synchronization among neuronal populations and their mean activity levels. *Neural Comput*. 11, 1389–411 (1999).
17. Rauschecker, J. P., Tian, B. & Hauser, M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114 (1995).
18. Zatorre, R. J., Evans, A. C. & Meyer, E. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci*. 14, 1908–1919 (1994).
19. Griffiths, T. D., Johnsrude, I., Dean, J. L. & Green, G. G. R. A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport* 18, 3825–3830 (1999).
20. Patterson, R. D., Handel, S., Yost, W. A. & Datta, A. J. The relative strength of the tone and the noise components in iterated rippled noise. *J. Acoust. Soc. Am*. 100, 3286–3294 (1996).
21. Duvernoy, H. M. *The Human Brain* (Springer, New York, 1999).