

## fMRI of the Central Auditory System

Deborah Ann Hall and Aspasia Eleni Paltoglou

### Abstract

Over the years, blood oxygen level-dependent (BOLD) fMRI has made important contributions to the understanding of central auditory processing in humans. Although there are significant technical challenges to overcome in the case of auditory fMRI, the unique methodological advantage of fMRI as an indicator of population neural activity lies in its spatial precision. It can be used to examine the neural basis of auditory representation at a number of spatial scales, from the micro-anatomical scale of population assemblies to the macro-anatomical scale of cortico-cortical circuits. The spatial resolution of fMRI is maximized in the case of mapping individual brain activity, and here it has been possible to demonstrate known organizational features of the auditory system that have hitherto been possible only using invasive electrophysiological recording methods. Frequency coding in the primary auditory cortex is one such example that we shall discuss in this chapter. Of course, noninvasive procedures for neuroscience are the ultimate aim and as the field moves towards this goal by recording in awake, behaving animals so human neuroimaging techniques will be increasingly relied upon to provide an interpretive link between animal neurophysiology at the multi-unit level and the operation of larger neuronal assemblies, as well as the mechanisms of auditory perception itself. For example, the neural effects of intentional behavior on stimulus-driven coding have been explored both in animals, using electrophysiological techniques, and in humans, using fMRI. While the feature-specific effects of selective attention are well established in the visual cortex, the effect of auditory attention in the auditory cortex has generally been examined at a very coarse spatial scale. Ongoing research in our laboratory has started to address this question and here we present preliminary evidence for frequency-specific effects of attentional enhancement in the human auditory cortex. We end with a brief discussion of several future directions for auditory fMRI research.

**Key words** Technical challenges, Frequency coding, Selective attention, Perceptual representation, Task specificity

---

### 1 Challenges of Auditory fMRI

The construction of a brain image using MR imaging depends upon the magnetic properties of hydrogen ions that, when placed in a static magnetic field, can absorb pulses of radiowave energy of a specific frequency. The time taken for the ion alignments to return to equilibrium after the radiofrequency (RF) pulse differs according to the surrounding tissue, thus providing the image contrast, for example between gray matter, white matter, cerebrospinal fluid, and bone. The use of MR techniques for detecting

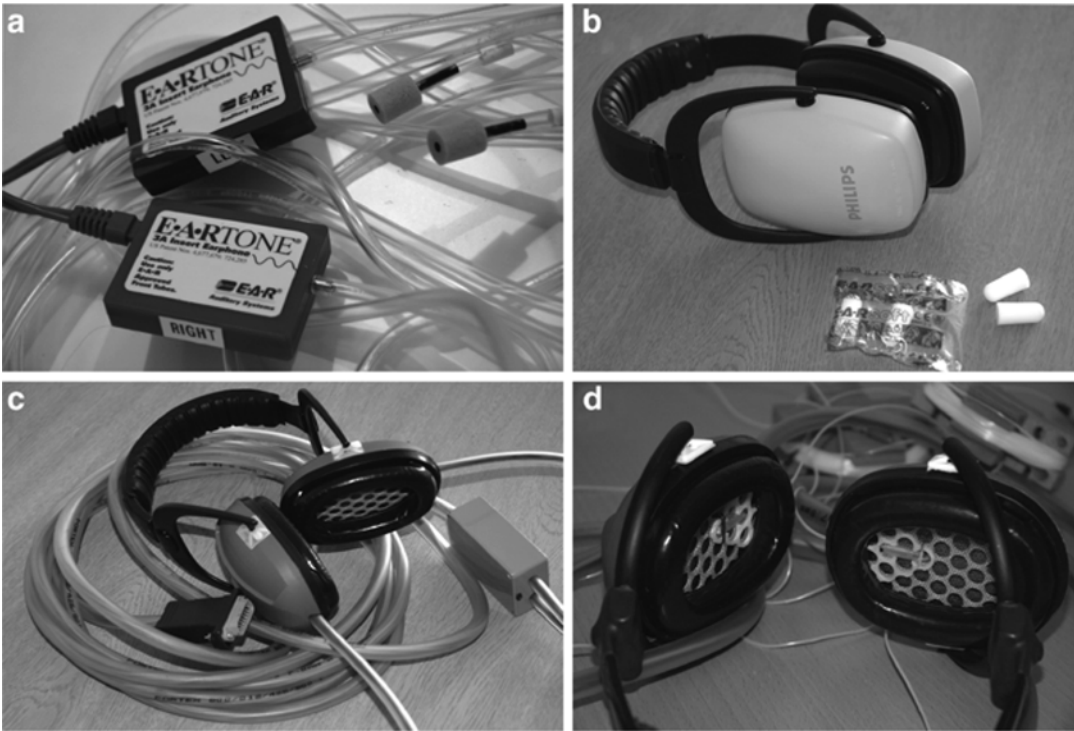
functional brain activation relies on two factors: first that local neural activity is a metabolically demanding process that is closely associated with a local increase in the supply of oxygenated blood to those active parts of the brain, and second that the different paramagnetic properties of oxygenated and deoxygenated blood produce measurable effects on the MR signal. The functional signal detected during fMRI is known as the blood oxygen level-dependent (BOLD) response. Essentially, the functional image represents the spatial distribution of blood oxygenation levels in the brain, and the small fluctuations in these levels over time are correlated with the stimulus input or cognitive task.

MR scanners operate using three different types of electromagnetic fields: a very high static field generated by a superconducting magnet, time-varying gradient magnetic fields, and pulsed RF fields. The latter two fields are much weaker than the first, but all pose a number of unique and considerable technical challenges for conducting auditory fMRI research within this hostile environment. In the first place, the static and time-varying magnetic fields preclude the use of many types of electronic sound presentation equipment, as well as preventing the safe scanning of patients who are wearing listening devices such as hearing aids or implants. Additionally, the high levels of scanner noise generated by the flexing of the gradient coils in the static magnetic field can potentially cause hearing difficulties. The scanner noise masks the perception of the acoustic stimuli presented to the subject in the scanner making it difficult to calibrate audible hearing levels and adding to the difficulty of the listening task. And finally, the scanner noise not only activates parts of the auditory brain, but also interacts with the patterns of activity evoked by experimental stimuli. Auditory fMRI poses a number of other challenges, not related to the hostile environment of the MR scanner, but related instead to the nature of the neural coding in the auditory cortex. The response of auditory cortical neurons to a particular class of sound is determined not only by the acoustic features of that sound, but also by its presentation context. For example, neurons respond strongly to the onset of sound events and thereafter tend to show rapid adaptation to that sound in terms of a reduction in their firing rate. Thus, the result of any particular auditory fMRI experiment will depend not only on the physical attributes of a stimulus, but also on the way in which the stimuli are presented. In this first section, we shall take each one of these issues in turn, introducing the problems in more detail as well as proposing some solutions.

## **1.1 Use of Electronic Equipment for Sound Presentation in the MR Scanner**

### **1.1.1 Problems**

The ideal requirement is a sound presentation system that produces a range of sound levels [up to 100-dB sound pressure level (SPL)], with low distortion, a flat frequency response, and a smooth and predictable phase response. The first commercially available solution utilized loudspeakers, placed away from the high static magnetic field, from which the sound was delivered through plastic tubes inserted into the ear canal (Fig. 1a) through a protective ear



**Fig. 1** MR compatible headsets for sound delivery and noise reduction: **(a)** tube phones system with foam ear inserts, **(b)** circum-aural ear defenders, plus foam ear plugs for passive noise reduction, **(c)** MRC IHR sound presentation headset combining commercially available electrostatic transducers in an industry standard ear defender, and **(d)** modified MRC IHR headset for sound presentation and for active noise cancellation (ANC), including an optical error microphone positioned underneath the ear defender

defender (Fig. 1b). One general disadvantage of the tube phone system is that the tubing distorts both the phase and amplitude of the acoustic signal, for example, by imposing a severe ripple on the spectra and reducing sound level, especially at higher frequencies. Another limitation is the leak of the scanner noise through the pipe walls to the pipe inner and hence the ear. Despite alternative systems now being readily available, tube phone systems are still commercially manufactured (e.g. Avotec Inc. Stuart, Florida, USA, [www.avotec.org/](http://www.avotec.org/)). The Avotec system has been specifically designed for fMRI use and boasts an equalizer to provide a reasonably flat audio output ( $\pm 5$  dB) across its nominal bandwidth (150 Hz–4.5 kHz) and a procedure for acoustic calibration that feeds a known electrical input signal to the audio system input and makes a direct acoustic output measurement at the headset.

Alternative electronic systems often used for psychoacoustical research deliver high-quality signals, but these systems are generally unsuitable for use in the MR environment because most headphones use an electromagnet to push and pull on a diaphragm to vibrate the air and generate sound. Of course, this electromagnet

is rendered inoperable by the magnetic fields in the MR scanner. Headphone components constructed from ferromagnetic material also disrupt the magnetic fields locally and induce signal loss or spatial distortion in areas close to the ears. In addition, the electronic components can be damaged by the static magnetic field, while electromagnetic interference generated by the equipment is detected by the MR receiver head coil. Electronic sound delivery systems for use in auditory fMRI research have been designed specifically to overcome these difficulties.

### 1.1.2 Solutions

Despite the restriction on the materials that can be used in a scanner, a number of different MR-compatible active headphone driving units have been produced. An ingenious system has been developed and marketed by one auditory neuroimaging research group (MR confon GmbH, Magdeburg, Germany, [www.mr-confon.de](http://www.mr-confon.de)). This system incorporates a unique, electrodynamic driver that uses the scanner's static magnetic field in place of the permanent magnets that are found in conventional headphones and loudspeakers. It produces a wide frequency range (less than 200 Hz–35 kHz) with a flat frequency response ( $\pm 6$  dB). Another company manufactures and supplies high-quality products for MRI, with a special focus on the fast-growing field of functional imaging (NordicNeuroLab AS, Bergen, Norway, [www.nordicneurolab.com/](http://www.nordicneurolab.com/)). Their audio system uses electrostatic transducers to ensure high performance. Electrostatic headphones generate sound using a conductive diaphragm placed next to a fixed conducting panel. A high voltage polarizes the fixed panel and the audio signal passing through the diaphragm rapidly switches between a positive and a negative signal, attracting or repelling it to the fixed panel and thus vibrating the air. Their technical specification claims a flat frequency response from 8 Hz to 35 kHz. The signal is transferred from the audio source to the headphones in the RF screened scanner room using either filters through a filter panel or fiber-optic cable through the waveguide.

Here at the MRC Institute of Hearing Research, we became engaged in auditory fMRI research well before such commercial systems were widely available and so, for our own purposes, we developed an MR-compatible headset (Fig. 1c) based on commercially available electrostatic headphones, modified to remove or replace their ferromagnetic components, and combined with standard industrial ear defenders to provide good acoustic isolation [1]. Our custom-built system delivers a flat frequency response ( $\pm 10$  dB) across the frequency range 50 Hz–10 kHz and has an output level capability up to 120-dB SPL. Again, the digital audio source, electronics, and power supply that drive the system are housed outside the RF screened scanner room to avoid electromagnetic interference with MR scanning, and all electrical signals passing into the screened scanner room are RF filtered.

## **1.2 Risk to Patients Who Are Wearing Listening Devices in the MR Scanner**

### *1.2.1 Problems*

No ferromagnetic components can be placed in the scanner bore as they would experience a strong attraction by the static magnetic field and potentially cause damage not only to the scanner and the listening device, but also to the patient. Induced currents in the electronics, caused directly by the time-varying gradient magnetic fields or the RF pulses, are an additional hazard to the electronic devices themselves, while some materials can also absorb the RF energy causing local tissue heating and even burns if in contact with soft tissue. For these reasons, there are restrictions on scanning people who have electronic listening devices. These include hearing aids, cochlear implants, and brainstem implants. Hearing aids amplify sound for people who have moderate to profound hearing loss. The aid is battery-operated and worn in or around the ear. Hearing aids are available in different shapes, sizes, and types, but they all work in a similar way. They all have a built-in microphone that picks up sound from the environment. These sounds are processed electronically and made louder, either by analogue circuits or digitally, and the resulting signals are passed to a receiver in the hearing aid where they are converted back into audible sounds. In contrast, cochlear and brainstem implants are both small, complex electronic devices that can help to provide a sense of sound to people who are profoundly deaf or severely hard-of-hearing. Cochlear implants bypass damaged portions of the inner ear (the cochlea) and directly stimulate the auditory nerve, while auditory brainstem implants bypass the vestibulocochlear nerve in cases when it is damaged by tumors or surgery and directly stimulate the lower part of the auditory brain (the cochlear nucleus). In general, both types of implant consist of an external portion that sits behind the ear and a second portion that is surgically placed under the skin. They contain a microphone, a sound processor (which converts sounds picked up by the microphone into an electrical code), a transmitter and receiver/stimulator (which receive signals from the processor and convert them into electric impulses), and finally an electrode array (which is a set of electrodes that collect the impulses from the stimulator and stimulate groups of auditory neurons). Coded information from the sound processor is delivered across the skin via electromagnetic induction to the implanted receiver/stimulator, which is surgically placed on a bone behind the ear.

### *1.2.2 Solutions*

Official approval for the manufacture of implant devices requires rigorous testing for susceptibility to electromagnetic fields, radiated electromagnetic fields, and electrical safety testing (including susceptibility to electrical discharge). However, such tests are conducted under normal conditions, not in the magnetic fields of an MR scanner. Some implant designs have been proven to be MR compatible [2–5], but they are not routinely supplied in clinical practice. Standard listening devices do not meet MR compatibility criteria and, for the patient, risks include movement of the device

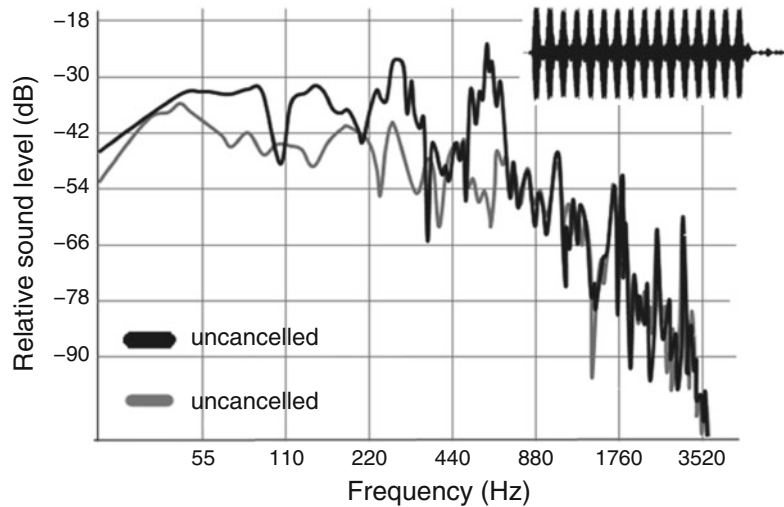
and localized heating of brain tissue, whereas, for the device, the electronic components may be damaged. Magnetic Resonance Safety Testing Services (MRSTS) is a highly experienced testing company that conducts comprehensive evaluations of implants, devices, objects, and materials in the MR environment (MRSTS, Los Angeles, CA, [www.magneticresonancesafetytesting.com/](http://www.magneticresonancesafetytesting.com/)). Testing includes approved assessment of magnetic field interactions, heating, induced electrical currents, and artifacts. A database of the devices and results of implant testing is accessible to the interested reader ([www.mrisafety.com/](http://www.mrisafety.com/)). However, auditory devices have generally been tested only at low magnetic fields (up to 1.5 T) because most clinical MR systems operate at this field strength. Since research systems typically operate at 3.0 T (for improved BOLD signal-to-noise ratio, BOLD SNR) it may be necessary for individual research teams to ensure the safety of their patients. For example, here at the MRC Institute of Hearing Research, we have recently assessed the risks of movement and localized tissue heating for two middle ear piston devices [6]. For the safety reasons discussed in this subsection, listeners who normally wear hearing aids could be scanned without their aid but, to compensate, have been presented with sounds amplified to an audible level. Given that implanted devices cannot be removed without surgical intervention, clinical imaging research of implantees has generally used other brain imaging methods, namely positron emission tomography [7].

### **1.3 Intense MR Scanner Noise and Its Effects on Hearing**

#### *1.3.1 Problems*

The scanning sequence used to measure the BOLD fMRI signal requires rapid on and off switching of electrical currents through the three gradient coils of wire in order to create time-varying magnetic fields that are required for selecting and encoding the three-dimensional image volume (in the  $x$ ,  $y$ , and  $z$  planes). This rapid switching in the static magnetic field induces bending and buckling of the gradient coils during MRI. As a result, the gradient coils act like a moving coil loudspeaker to produce a compression wave in the air, which is heard as acoustic noise during the image acquisition. Scanner noise increases nonlinearly with static magnetic field strength, such that ramping from 0.5 to 2 T could account for a rise in sound level of as much as 11-dB SPL [8]. A brain scan is composed of a set of two-dimensional “slices” through the brain. Gradient switching is required for each slice acquisition and so an intense scanner “ping” occurs each time a brain slice is collected. Each ping lasts about 50 ms and so during fMRI, each scan is audible as a rapid sequence of such “pings” (see inset in Fig. 2 for an example of the amplitude envelope of the scanner noise).

The dominant components of the noise spectrum are composed of a peak of sound energy at the gradient switching frequency plus its higher harmonics. Most of the energy lies below 3 kHz. Secondary acoustic noise can be produced if the vibration of the

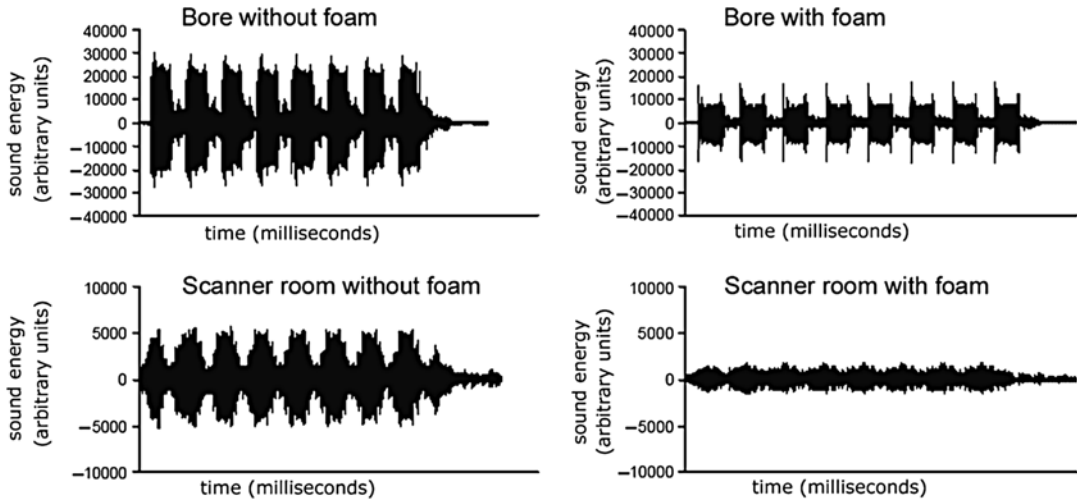


**Fig. 2** Typical frequency spectrum of the scanner noise generated during blood oxygen level-dependent (BOLD) fMRI. This example was measured in the bore of a Philips Intera 3.0 Tesla scanner. The *black line* (uncancelled) indicates the acoustic energy of the noise recorded under normal scanning conditions. The *gray line* (canceled) indicates the residual acoustic energy at the ear when the active noise cancellation (ANC) system is operative. The *inset (upper right)* shows an example of the amplitude envelope of the scanner noise for a brain scan consisting of 16 slices corresponding to a sequence of 16 intense “pings”

coils and the core on which they are wound conducts through the core supports to the rest of the scanner structure. These secondary noise characteristics depend more on the mechanical resonances of the coil assemblies than on the type of imaging sequence and they tend to be the dominant contributor to the bandwidth and the spectral envelope of the noise. In this example of the frequency spectrum captured from a BOLD fMRI scanning sequence that was run on a Philips 3 Tesla Intera (Fig. 2), the spectrum has a peak component at 600 Hz with several other prominent pseudo-harmonics at 300, 1080, and 1720 Hz. The sound level measured in the bore of the scanner is typically 99-dB SPL [98 dB(A) using an A-weighting], measured using the maximum “fast” root-mean-square (RMS) time constant (125 ms). Clearly, exposure to such an intense sound levels without protection is likely to cause a temporary threshold shift in hearing and tinnitus, and it could be permanently damaging over a prolonged dosage [9].

### 1.3.2 Solutions

The simplest way to treat the intense noise is to use ear protection in the form of ear defenders and/or ear plugs (shown in Fig. 1b). Foam ear plugs can compromise the acoustic quality of the experimental sounds delivered to the subject and so ear defenders are preferable.



**Fig. 3** Acoustic waveforms of the scanner noise measured with and without a lining of acoustic damping foam in the bore of the scanner. Our data demonstrate that the foam reduces the sound pressure level (SPL) at the position of the subject's head and in scanner room by a significant margin (about 8 dB). The segment of scanner noise that is illustrated here has a duration of approximately 1 s

Typically, transducers are fitted into sound attenuating earmuffs to reduce the ambient noise level at the subject's ears. Attenuation of the external sound by up to 40 dB can be achieved in this manner, although the level of reduction drops off at the high-frequency end of the spectrum. Commercial sound delivery systems all incorporate passive noise attenuation of this sort. An additional method of noise reduction is to line the bore of the scanner with a sound-energy absorbing material ([10]; see also [www.ihr.mrc.ac.uk/research/technical/soundsystem/](http://www.ihr.mrc.ac.uk/research/technical/soundsystem/)). The results of a set of measurements directly comparing the sound intensity of the scanner noise with and without the foam lining are shown in Fig. 3. However, this strategy does not provide a feasible solution because neither the design of the scanner bore nor the automated patient table are suited to the permanent installation of a foam lining and some types of acoustic foam can present risks of noxious fumes if they catch fire.

Some manufacturers have attempted to minimize scanner sound levels by modifying the design of the scanner hardware. For example, MR scanners manufactured by Toshiba (Toshiba America Medical systems, Inc., [www.medical.toshiba.com/](http://www.medical.toshiba.com/)) incorporate Pianissimo technology—employing a solid foundation for gradient support, integrating sound dampening material in the gradient coils and enclosing them in a vacuum to reduce acoustic noise, even at full gradient power. This technology claims to reduce scanner noise by up to 90% [11]. Subjects are reported to hear sounds at the volume of gentle drumming instead of the jackhammer noise level of other MR systems.



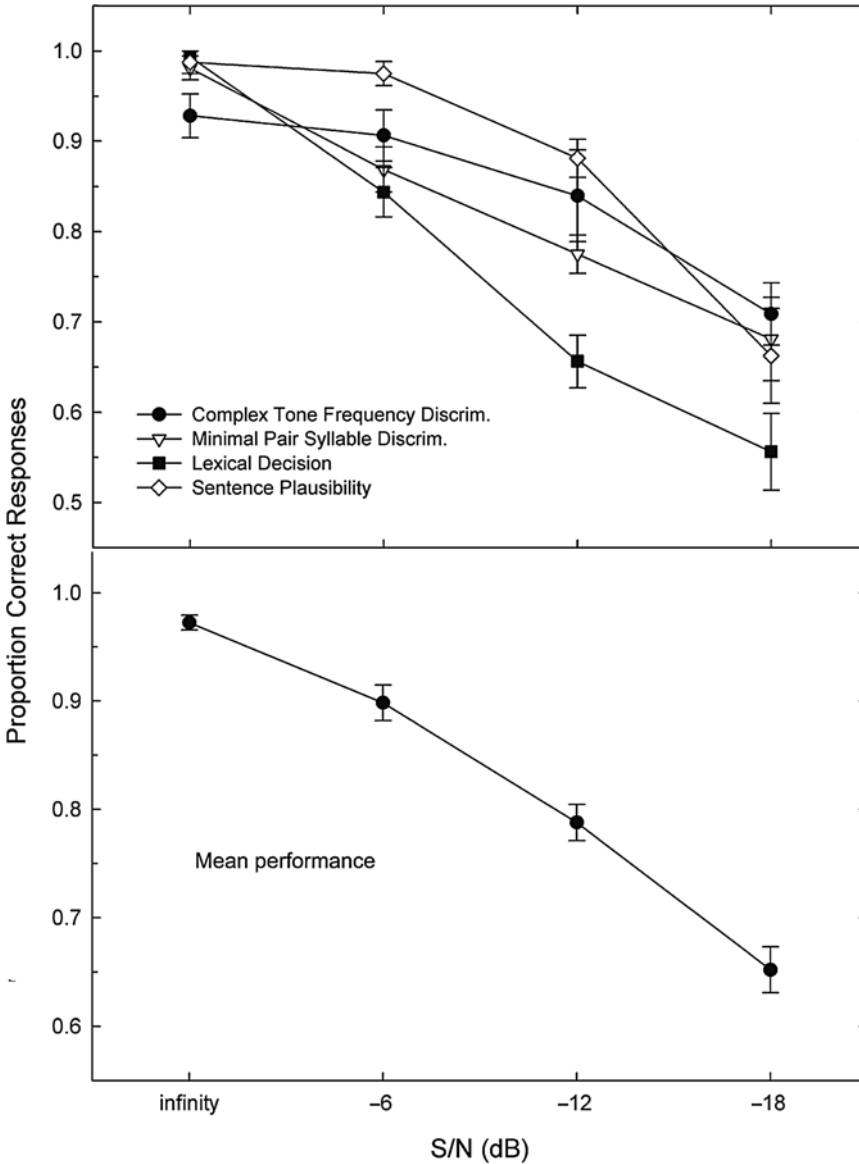
Another solution is to run modified pulse sequences that reduce acoustic noise by slowing down the gradient switching. This approach is based on the premise that the spectrum of the acoustic noise is determined by the product of the frequency spectrum of the gradient waveforms and the frequency response function of the gradient system [12]. The frequency response function is generally substantially reduced at low frequencies (i.e. below 200 Hz) and so the sound level can be reduced by using gradient pulse sequences whose spectra are band limited to this low-frequency range using pulse shapes with smooth onset and offset ramps [13]. A low-noise fast low-angle shot (FLASH) sequence can be modified to have a long gradient ramp time (6000  $\mu$ s) and it generates a peak sound level of 48-dB SPL measured at the position of the ear. This type of sequence has been used for mapping central auditory function [14]. However, the low noise is achieved at the expense of slower gradient switching, extending the acquisition time. Low-noise sequences are not suitable for rapid BOLD imaging in which the fundamental frequency of the gradient waveform is greater than 200 Hz.

## **1.4 The Effect of Scanner Noise on Stimulus Audibility**

### **1.4.1 Problems**

Not only is the intense scanner noise a risk for hearing, but it also masks the perception of the acoustic stimuli presented to the subject. The exact specification of the acoustic signal-to-scanner-noise ratio (acoustic SNR) in fMRI studies using auditory stimuli is a potentially complicated matter. Nevertheless, we have sought to establish the relative difference between the stimulus level and the scanner noise level at the ear, by measuring these signals using a reference microphone placed inside the cup of the ear defender while participants perform a signal detection in noise task. Detection thresholds for a narrow band noise centered at the peak frequency of the scanner noise (600 Hz) are elevated when the target coincides with the scanner noise. We have demonstrated an average 11-dB shift in the 71% detection threshold for the 600-Hz target when we modulate the perceived level of the scanner noise using active noise cancellation (ANC) methods (see later).

This evidence suggests that even with hearing protection, whenever the scanner noise coincides with the presented sound stimulus it produces changes in task performance and probably also increases the attentional demands of the listening task. The frequency range of the scanner acoustic noise is crucial for speech intelligibility, and speech experiments can be particularly compromised by a noisy environment ([15]; for review, see [16]). A recent study has quantified the effect of acoustic SNR using four listening tasks: pitch discrimination of complex tones, same/different judgments of minimal-pair nonsense syllables, lexical decision, and judgement of sentence plausibility [17]. Across these tasks, performance was assessed in silence (acoustic SNR=infinity) and in a background of MR scanner noise at the three acoustic SNR levels



**Fig. 4** Mean performance in a simulated scanning environment across four acoustic signal-to-noise ratios [17]. The *top panel* plots the proportion of correct responses on the individual tasks, while the *bottom panel* shows the overall mean performance (SNR signal-to-noise ratio, dB decibels)

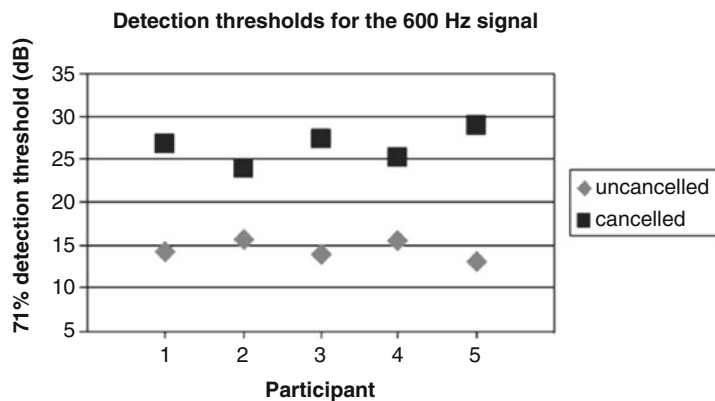
(-6, -12, and -18 dB). Performance of normally hearing listeners significantly decreased as a function of the noise (Fig. 4). Even at -6 dB acoustic SNR, participants made many more errors than in quiet listening conditions ( $p < 0.01$ ). Thus, across a range of auditory tasks that vary in linguistic complexity, listeners are highly susceptible to the disruptive impact of the intense noise associated with fMRI scanning.

## 1.4.2 Solutions

The aggregate noise dosage can be reduced by acquiring either a single or at least very few brain slices, but at the expense of only a partial view of brain activity [18]. For whole brain fMRI, other strategies are required.

One novel method that has been developed and evaluated at our Institute combines optical microphone technology with an active noise controller for significant attenuation of ambient noise received at the ears [19]. The canceller is based upon a variation of the single channel feed-forward filtered-x adaptive controller and uses a digital signal processor to achieve the noise reduction in real time. The canceler minimizes the noise pressure level at a specific control point in space that is defined by the position of the error microphone, positioned underneath the circum-aural ear defender of the headset (*see* Fig. 1d). In 2001, we published a psychophysical assessment of the system using a prototype system built in the laboratory that utilized a loudspeaker as the noise generator [19]. This system produced 10–20 dB of subjective noise reduction between 250 Hz and 1 kHz and smaller amounts at higher frequencies. More recently, we have obtained psychophysical threshold data in a Philips 3 Tesla scanner confirming that the same level of cancellation is achieved in the real scanner environment (Fig. 5; [20]). Again, the subjective impression of the scanner noise is the volume of gentle drumming when the sound system is operating in its canceled mode. Thus, it is possible to achieve a high level of noise attenuation by combining both passive and active methods.

A much more common strategy for reducing the masking influence of the concomitant scanner noise combines a passive method of ear protection with an experimental protocol that carefully controls



**Fig. 5** Performance on a signal detection in noise task measured in a real scanning environment [Philips 3.0 Tesla MR scanner during blood oxygen level-dependent (BOLD) fMRI]. The data show that when the noise canceller was operative, the sound level of the signal could be 8–16 dB softer (depending upon the listener) in order to achieve the same detection performance

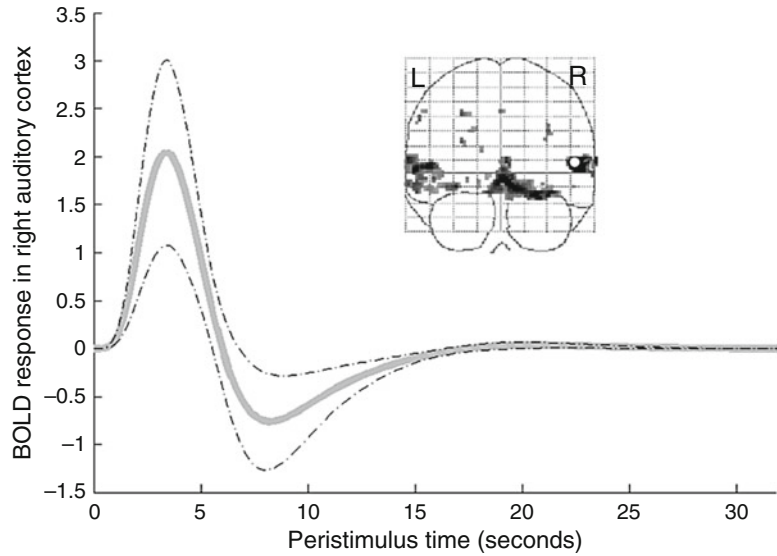
the timing between stimulus presentation and image acquisition so that sound stimuli can be delivered during brief periods of quiet in between successive brain scans [21]. Specific details of several pulse sequence protocols that reduce the masking effects of scanner noise are discussed in more detail in the next subsection.

## **1.5 The Effect of Scanner Noise on Sound-Related Activation in the Brain**

### **1.5.1 Problems**

To increase the BOLD SNR, it is necessary to acquire a large number of scans in each condition in an fMRI experiment. Typically, an experimenter would collect many hundreds of brain scans in a single session, with the time in between each scan chosen to be as short as the scanner hardware and software will permit. Remember that, for fMRI, an intense “ping” is generated for each slice of the scan and so of course this means that the participant can easily be subjected to several thousand repeated “pings” of noise during the experiment. Not only does this scanner noise acoustically mask the presented sound stimuli, but the elevated baseline of sound-evoked activation due to the ambient scanner noise also makes the experimentally induced auditory activation more difficult to detect statistically. Much of the work examining the influence of acoustic scanner noise has been directed toward its capacity to interfere with the study of audition or speech perception by producing activation of various brain regions, especially the auditory cortex [22–25]. Several studies highlight the reduced activation signal (i.e. the difference between stimulation and baseline conditions) in the auditory cortex when the amount of prior scanner noise is increased, demonstrating that the scanner noise effectively masks the detection of auditory activation [22, 26, 27]. In another example, taken from one of the early fMRI experiments conducted at the MRC Institute of Hearing Research, we used a specially tailored scanning protocol to measure the amplitude and the time course of the BOLD response to a high-quality recording of a single burst of scanner noise presented to participants over headphones [24]. Our results revealed a reliable transient increase in the BOLD signal across a large part of the auditory cortex. As in many other brain regions, the evoked response to this single brief stimulus event was smoothed and delayed in time. It rose to a peak by 4–5 s after stimulus onset and decayed by 5–8 s after stimulus offset [24]. Its amplitude reached about 1.5% of the overall signal change, which is considerable considering that stimulus-related activation usually accounts for a BOLD signal change of approximately 2–5%. Figure 6 illustrates the canonical BOLD response to a noise onset.

In many fMRI experimental paradigms, regions of stimulus-evoked activation are detected by comparing the BOLD scans acquired during one sound condition with the BOLD scans acquired during another condition, which could be either a condition in which a different type of sound was presented or no sound (known as a baseline “silent” condition) was presented. Activation is defined as those parts of the brain that demonstrate a statistically significant



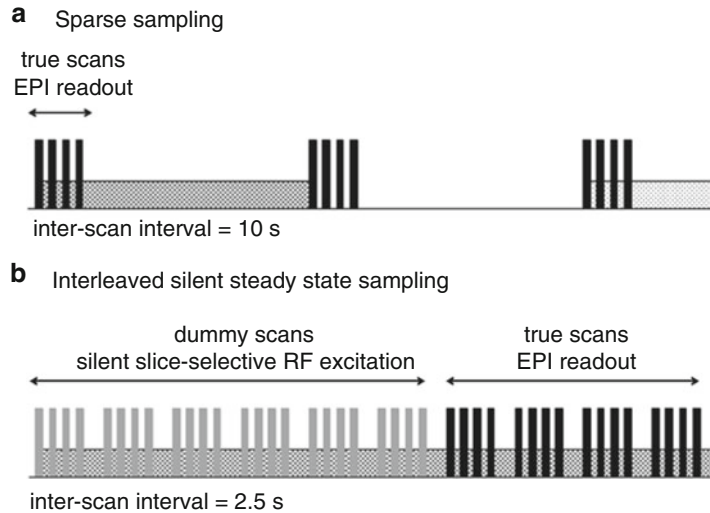
**Fig. 6** Transient blood oxygen level-dependent (BOLD) response to a noise onset. The graph shows the fitted response and the 90% confidence interval. This example illustrates all the characteristic features of the transient BOLD response—a peak at 4-s post-stimulus onset followed by an undershoot and then return to baseline at 16 s

difference between the two conditions. For example, let us consider the simplest case in which one condition contains a sound and the other does not. Since the scanner noise is present throughout, the sound condition effectively contains both stimulus and scanner noise, while the baseline condition also contains the scanner noise (i.e. it is not silent). Given the spectrotemporal characteristics of the scanner noise, it generates widespread sound-related activity across the auditory cortex. Thus, the subtraction analysis for detecting activation is sensitive only to whatever is the small additional contribution of the sound stimulus to auditory neural activity.

### 1.5.2 Solutions

A number of different scanning protocols have been used to minimize the effect of the scanner acoustic noise on the measured patterns of auditory cortical activation. In this section, we will describe two of these, but before we do, we need to consider some important details about the time course of the BOLD response to the scanner noise and introduce some new terms.

During an fMRI experiment, the BOLD response to the scanner noise spans two different temporal scales. First, the “ping” generated by the acquisition of one slice early in the scan may induce a BOLD response in a slice, which is acquired later in the same scan if that later scan is positioned over the auditory cortex. We shall call this inter-slice interference. Inter-slice interference is maximally reduced when all slices in the scan are acquired in rapid



**Fig. 7** Two scanning protocols that have been used to minimize the effect of the scanner acoustic noise on the measured patterns of auditory cortical activation. See text for further explanation (*s* seconds, *EPI* echo-planar imaging, *RF* radiofrequency)

succession and the total duration of the scan is not more than 2 s [26]. A common term for the scanning protocol that uses a minimum inter-slice interval is a clustered-acquisition sequence. Edmister et al. [28] found that the clustered-acquisition sequence provides an advantageous auditory BOLD SNR compared with a conventional scanning protocol. The second form of interference is called inter-scan interference. This occurs when the scanner noise evokes an auditory BOLD response that extends across time to subsequent scans, predominantly when the interval between scans is as short as the MR system will permit. Reducing the inter-scan interference can easily be achieved by extending the period between scans (the inter-scan interval). By separately manipulating the timing between slices and between scans, we can reduce the inter-slice and inter-scan interference independently of one another. When the clustered-acquisition sequence is combined with a long (e.g. 10 s) inter-scan interval, the activation associated with the experimental sound can be separated from the activation associated with the scanner sound (Fig. 7a). Furthermore, because the scanner sound is temporally offset, it does not produce acoustical masking and does not distract the listener. This scanning protocol is commonly known as sparse sampling [21]. Sparse sampling is often the scanning protocol of choice for identifying auditory cortical evoked responses in the absence of scanner noise (see e.g. [29–33]). However, it requires a scanning session that is longer than that of conventional “continuous” protocols in order to acquire the same amount of imaging data, and participants can be intolerant of long

sessions. It also relies upon certain assumptions about the time to peak of the BOLD response after stimulus onset and a sustained plateau of evoked activity for the duration of the stimulus.

A second type of scanning protocol acquires a rapid set of scans following each silent period in order to avoid some of the aforementioned difficulties—“interleaved silent steady state” sampling [34]. The increased number of scans permits a greater proportion of scanning time to be used for data acquisition and at least partial mapping of the time course of the BOLD response (Fig. 7b). However, some pulse programming is required to avoid T1-related signal decay during the data acquisition, hence ensuring that signal contrast is constant across successive scans. The software modification maintains the longitudinal magnetization in a steady state throughout the scanning session by applying a train of slice-selective excitation pulses (quiet dummy scans) during each silent period.

## **1.6 The Effect of Stimulus Context: Neural Adaptation to Sounds**

### *1.6.1 Problems*

The acoustic environment is typically composed of one or more sound sources that change over time. Over the years, both psychophysical and electrophysiological studies have amply demonstrated that stimulus context strongly influences the perception and neural coding of individual sounds, especially in the context of stream segregation and grouping [35–37]. A simple example of the influence of stimulus context is forward masking, which occurs when the presence of one sound increases the detection threshold for the subsequent sound. The perceptual effects of forward masking are strongest when the spectral content of the first sound is similar to the second sound, when there is no delay between the two sounds, and when the masker duration is long [38]. Forward inhibition typically lasts from 70 to 200 ms. This type of suppression has not only been demonstrated in anesthetized preparations, but also in awake primates. In the latter case, suppression was seen to extend up to 1 s in time [39]. As well as tone–tone interactions, neural firing rate is sensitive to stimulus duration. Neurons respond strongly to the onset of a sound and their response decays thereafter. Many illustrative examples can be found in the literature, especially in cases where longer duration sounds are presented (e.g. 750–1500 ms in the case of Bartlett and Wang [38], see their Fig. 4).

By transporting these well-established paradigms into a neuroimaging experiment, researchers are beginning to address the context dependency of neural coding in humans. One way in which the effect of sound context on the auditory BOLD fMRI signal has been examined is in terms of different repetition rates [19, 40]. This is conceptually analogous to the presentation rate manipulations of the forward masking studies described earlier, but goes beyond the simple case of two-tone interactions. In the fMRI studies, stimuli were long trains of noise bursts presented at different rates. The slowest rate was 2 Hz and the fastest rate was 35 Hz, with intermediate rates being 10 and 20 Hz. Noise bursts at each

repetition rate were presented in prolonged blocks of 30 s, each followed by a 30-s “silent” period. During sound presentation, scans were acquired at a short inter-scan interval (approximately 2 s) so that the experimenters could reconstruct the 30-s time course of the BOLD response to each of the different repetition rates, hence determining the multi-second time pattern of neural activity. The scans were positioned so that a number of different auditory sites in the ascending auditory system could be measured: (1) the inferior colliculus in the midbrain, (2) the medial geniculate nucleus in the thalamus, and (3) Heschl’s gyrus and the superior temporal gyrus in the cortex. The plots of the BOLD time course demonstrated a systematic change in its shape from midbrain up to cortex. In the inferior colliculus, the amplitude of the BOLD response increased as a function of repetition rate while its shape was sustained throughout the 30-s stimulus period. In the medial geniculate body, increasing rate also produced an increase in BOLD amplitude with a moderate peak in the BOLD shape just after stimulus onset. Repetition rate exerted its largest effect in the auditory cortex. The most striking change was in the shape of the BOLD response. The low repetition rate (2 Hz) elicited a sustained response, whereas the high rate (35 Hz) elicited a phasic response with prominent peaks just after stimulus onset and offset. The follow-up study [40] confirmed that it was the temporal envelope characteristics of the acoustic stimulus, not its sound level or bandwidth, that strongly influenced the shape of the BOLD response. The authors offer a perceptual interpretation of the neural response to different repetition rates. The shift in the shape of the cortical BOLD response from sustained to phasic corresponds to a shift from a stimulus in which component noise bursts are perceptually distinct to one in which successive noise bursts fuse to become individually indistinguishable. The onset and offset responses of the phasic response coincide with the onset and offset of a distinct, meaningful event. The logical conclusion to this argument is that the succession of individual perceptual events in the low repetition rate conditions defines the sustained BOLD response observed at the 2-Hz rate. It is clear from these results that while the amplitude of the BOLD response to sound can inform us about the tuning properties of the underlying neural population (e.g. sensitivity to repetition rate), other properties of the BOLD response, such as its shape, provide different information about neural coding (e.g. segmentation of the auditory environment into perceptual events).

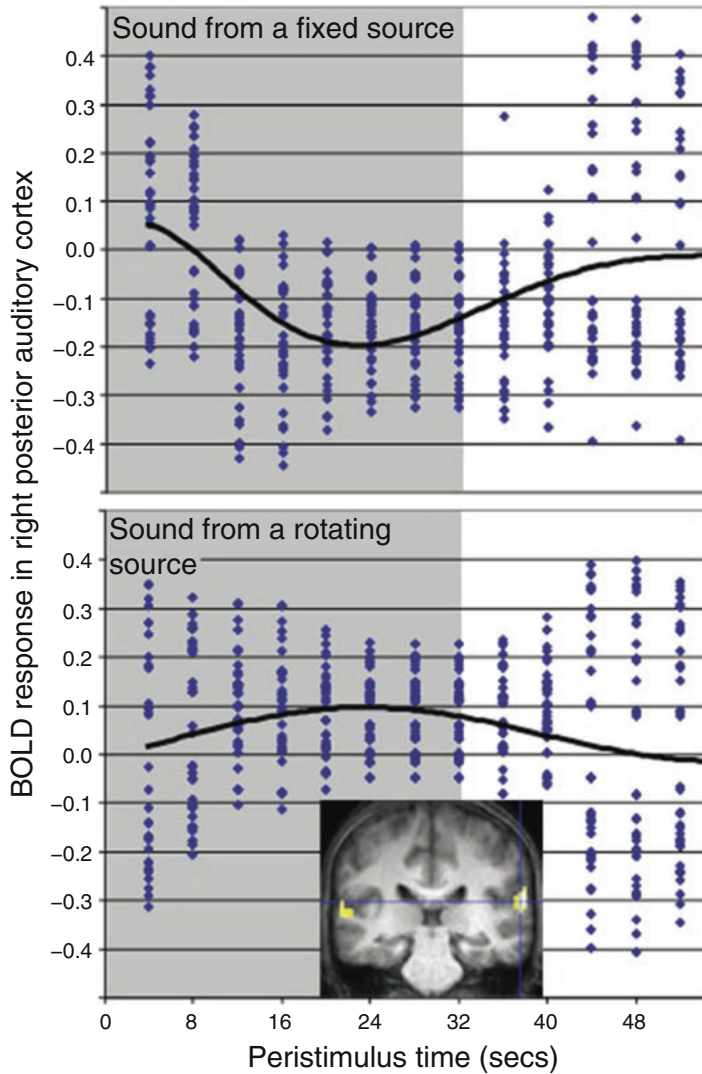
It is crucial that these contextual influences on the BOLD signal are accounted for in the design and/or interpretation of auditory fMRI experiments. To illustrate this case in point, I use a set of our own experimental data [41]. In this experiment, one of the sound conditions was a diotic noise (identical signal at the two ears) presented continuously for 32 s at a constant sound level



(~86-dB SPL) and at a fixed location in the azimuthal plane. Scans were acquired every 4 s throughout the stimulus period. When the scans acquired during this sound condition were combined together and contrasted against the scans acquired during the “silent” baseline condition, no overall significant activation was obtained ( $p > 0.001$ ). We interpret this lack of activation as evidence that the auditory response had rapidly habituated to a static signal. This conclusion is confirmed by plotting out the time course of the response at one location within the auditory cortex. The initial transient rise in the BOLD response at the onset of the sound begins to decay at about 4 s and this reduction continues across the stimulus epoch. The end of the epoch is characterized by a further rise in the BOLD response, elicited by the other types of sound stimuli that were presented in the experiment (Fig. 8a).

### 1.6.2 Solutions

It is common for auditory fMRI experiments to use a blocked design in which a sound condition is presented over a prolonged time period that extends over many seconds, even tens of seconds. Indeed as we described in Sect. 1.5, the blocked design is at the core of the sparse sampling protocol, and so the risk of neural adaptation is a legitimate one. The BOLD signal detection problem caused by neural adaptation is often circumvented by presenting the stimulus of interest as a train of stimulus bursts at a repetition rate that elicits the sustained cortical response (e.g. 2 Hz). Many of the auditory fMRI experiments that have been conducted over the years in our research group have taken this form [30, 31, 42–44]. Alternatively, if the stimulus contains dynamic spectrotemporal changes, then it is not always necessary to pulse the stimulus on and off. To illustrate this case in point, I return to a set of our own experimental data [41]. In this experiment, one of the sound conditions was a broadband noise convolved with a generic head-related transfer function to give the perceptual impression of a sound source that was continuously rotating around the azimuthal plane of the listener. Although the sound was presented continuously for 32 s, the filter functions of the pinnae imposed a changing frequency spectrum and the head shadow effect imposed low-rate amplitude modulations in the sound envelope presented to each ear. When the scans acquired during this sound condition were combined together and contrasted against the scans acquired during the “fixed sound source” condition, widespread activation was obtained ( $p < 0.001$ ) across the posterior auditory cortex (planum temporale): an area traditionally linked with spatial acoustic analysis. The time course of activation demonstrates a sustained BOLD response across the entire duration of the epoch (Fig. 8b). The sustained response contrasts with the transient response observed for the fixed sound source condition (Fig. 8a).



**Fig. 8** Adjusted blood oxygen level-dependent (BOLD) response (measured in arbitrary units) across the 32-s stimulus epoch shaded in *gray* (a) for a sound from a fixed source and (b) for a sound from a rotating source. Adjusted values are combined for all six participants and the trend line is indicated using a polynomial sixth order function. The response for both stimulus types is plotted using the same voxel location in the planum temporale region of the right auditory cortex (coordinates  $x$  63,  $y$  -30,  $z$  15 mm). The position of this voxel is shown in the *inserted panel*. The activation illustrated in this *insert* represents the subtraction of the fixed sound location from the rotating sound conditions ( $p < 0.001$ )

## 2 Examples of Auditory Feature Processing

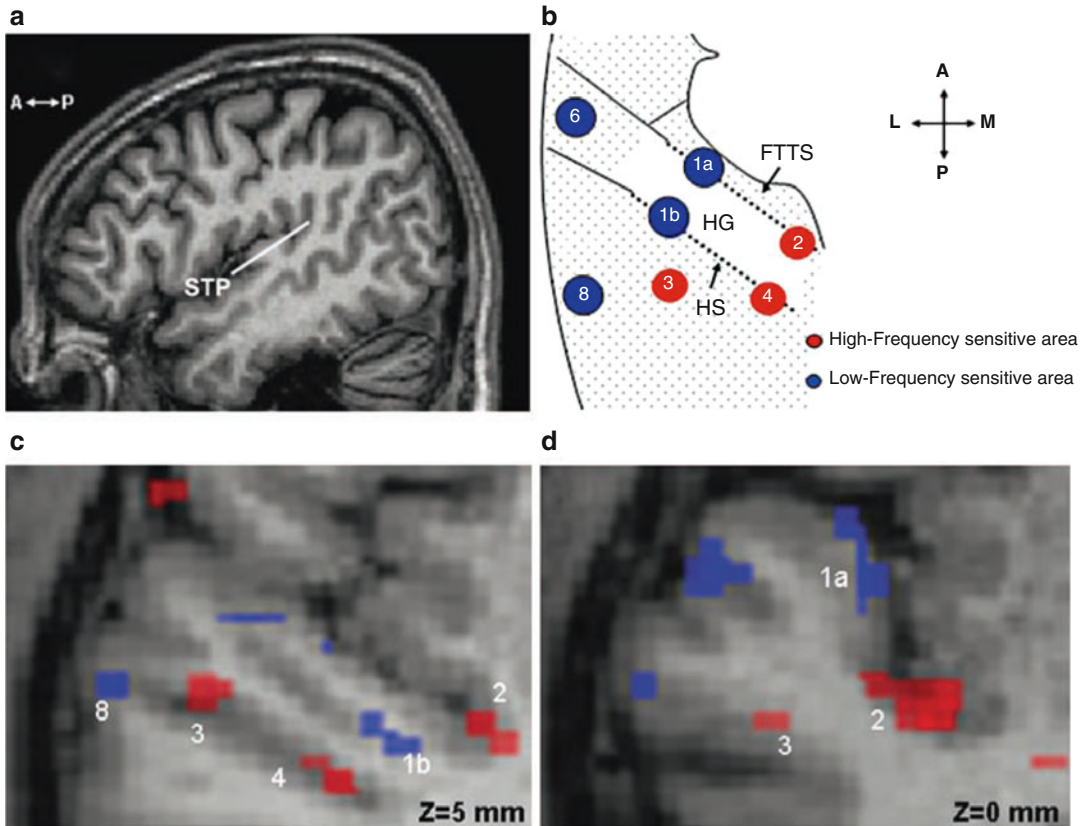
### 2.1 The Representation of Frequency in the Auditory Cortex

Within the inner ear, an incoming sound is separated into its individual frequency components by the way in which the energy at different frequencies travels along the cochlear partition [45]. High-frequency tones maximally stimulate those nerve fibers near the base of the cochlea while low-frequency tones are best coded

towards the apex. This cochleotopic representation persists throughout the auditory pathway where it is referred to as a tonotopic map. Within the mammalian auditory cortex, electrophysiological recordings have revealed many tonotopic maps [46, 47]. Within each map, neurons tuned to the same sound frequency are colocalized in a strip across the cortical surface, with an orderly progression of frequency tuning across adjacent strips. Frequency tuning is sharper in the primary auditory fields than it is in the surrounding nonprimary fields, and so the most complete representations of the audible frequency range are found in the primary fields. Primates have at least two tonotopic maps in primary auditory cortex, adjacent to one another and with mirror-reversed frequency axes. It is possible to demonstrate tonotopy by fMRI as well as by electrophysiology, even though frequency selectivity deteriorates at the moderate to high sound intensities required for fMRI sound presentation. As a recent example, mirror-symmetric frequency gradients have been confirmed across primary auditory fields using high-resolution fMRI at 4.7 T in anesthetized macaques and at 7.0 T in awake behaving macaques [48]. This section describes results from several fMRI experiments that have sought to demonstrate tonotopy in the human auditory cortex.

fMRI is an ideal tool for exploring the spatial distribution of the frequency-dependent responses across the human auditory cortex because it provides good spatial resolution and the analysis requires few a priori modeling assumptions (*see* [49] for a review). In addition, it is possible to detect statistically significant activation using individual fMRI analysis. This is important when determining fine-grained spatial organization because averaging data across different listeners would inevitably blur the subtle distinctions. A number of recent studies have sought to determine the organization of human tonotopy [29, 33, 50–52]. To avoid the problem of neural adaptation discussed in Sect. 1.6, experimenters chose stimuli that would elicit robust auditory cortical activation. For example, Talavage et al. [51, 52] presented amplitude-modulated signals, while Schönwiesner et al. [50] presented sine tones that were frequency modulated across a narrow bandwidth. Langers et al. [33] used a signal detection task in which the tone targets at each frequency were briefly presented (0.5 s). In agreement with the primate literature, evidence for the presence of tonotopic organization is at its most apparent within the primary auditory cortex while frequency preferences in the surrounding nonprimary areas are more erratic [33]. Thus, we shall consider in more detail the precise arrangement of tonotopy in the primary region.

In their first study, Talavage et al. [51] contrasted pairs of low (<66 Hz) and high (>2490 Hz) frequency stimuli of moderate intensity and sufficient spectral separation to produce spatially resolvable differences in activation (low>high and high>low) across the auditory cortical surface. These activation foci were considered to define the endpoints of a frequency gradient. In total,



**Fig. 9** (a) Sagittal view of the brain with the *oblique white line* denoting the approximate location and orientation of the schematic view shown in panel (b) along the supratemporal plane. (b) Schematic representation of the most consistently found high (*red*) and low (*blue*) frequency-sensitive areas across the human auditory cortex reported by Talavage et al. [50, 51]. The primary area is shown in *white* and the nonprimary areas are shown by *dotted shading*. Panels (c) and (d) illustrate the high- (*red*) and low- (*blue*) frequency sensitive areas across the left auditory cortex of one participant (unpublished data). Two planes in the superior-inferior dimension are shown ( $z=5$  mm and  $z=0$  mm above the CA-CP line). A anterior, P posterior, M medial, L lateral, HG Heschl's gyrus, HS Heschl's sulcus, FTTS first transverse temporal sulcus, STP supratemporal plane

Talavage et al. identified eight frequency-sensitive sites across Heschl's gyrus (HG, the primary auditory cortex) and the surrounding superior temporal plane (STP, the nonprimary auditory cortex). Each site was reliably identified across listeners and the sites were defined by a numerical label [1–8].

Foci 1–4 occurred around the medial two-thirds of HG and are good candidates for representing frequency coding within the primary auditory cortex (Fig. 9). Finding several endpoints does not provide direct confirmation of tonotopy because tonotopy necessitates a linear gradient of frequency sensitivity. Nevertheless, Talavage et al. argued that the foci 1–3 were at least consistent with predictions from primate electrophysiology.

The arrangement of the three foci encompassed the primary auditory cortex, suggested a common low-frequency border, and had a mirror-image reversed pattern. This interpretation was criticized by Schönwiesner et al. [50] who stated that it was wrong to associate these foci with specific tonotopic fields because pairs of low- and high-frequency foci could not clearly be attributed to specific frequency axes nor to anatomically defined fields. Indeed, in their own study, Schönwiesner et al. [50] did not observe the predicted gradual decrease in frequency-response amplitude at locations away from the best-frequency focus, but instead found a rather complex distribution of response profiles. Their explanation for this finding was that the regions of frequency sensitivity reflected not tonotopy, but distinct cortical areas that each preferred different acoustic features associated with a limited bandwidth signal.

Increasing the BOLD SNR might be necessary for characterizing some of the more subtle changes in the response away from best frequency and more recent evidence using more sophisticated scanning techniques does support the tonotopy viewpoint. Frequency sensitivity in the primary auditory cortex was studied using a 7-T ultra-high field MR scanner to improve the BOLD SNR and to provide reasonably fine-grained ( $1 \text{ mm}^3$ ) spatial resolution [29]. Formisano et al. [29] sought to map the progression of activation as a smooth function of tone frequency across HG. Frequency sensitivity was mapped by computing the locations of the best response to single frequency tones presented at a range of frequencies (0.3, 0.5, 0.8, 1, 2, and 3 kHz). Flattened cortical maps of best frequency revealed two mirror-symmetric gradients (high-to-low and low-to-high) traveling along HG from an anterolateral point to the posteromedial extremity. In general, the amplitude of the BOLD response decreased as the stimulating tone frequency moved away from the best frequency tuning characteristics of the voxel. A receiver coil placed close to the scalp over the position of the auditory cortex is another way to achieve a good BOLD SNR and this was the method used by Talavage et al. [52]. Talavage et al. measured best-frequency responses to an acoustic signal that was slowly modulated in frequency across the range 0.1–8 kHz. Again, the results confirmed the presence of two mirror-symmetric maps that crossed HG (extending from the anterior first transverse temporal sulcus to the posterior Heschl's sulcus) and shared a low-frequency border.

Although more evidence will be required before a clear consensus is established, the studies presented in this section have made influential contributions to the understanding of frequency representation in the human auditory cortex and its correspondence to primate models of auditory coding.

## **2.2 The Influence of Selective Attention on Frequency Representations in Human Auditory Cortex**

We live in a complex sound environment in which many different overlapping auditory sources contribute to the incoming acoustical signal. Our brains have a limited processing capacity and so one of the most important functions of neural coding is to separate out these competing sources of information. One way to achieve this is by filtering out the uninformative signals (the “ground”) and attending to the signal of interest (the “figure”). Competition between incoming signals can be resolved by a bottom-up, stimulus-driven process (such as a highly salient stimulus that evokes an involuntary orienting response), or it can be resolved by a top-down, goal-directed process (such as selective attention). Selective attention provides a modulatory influence that enables a listener to focus on the figure and to filter out or attenuate the ground [53].

Visual scientists have shown that attention can be directed to the features of the figure (feature-based attention, for a review *see* [54]) or to the entire figure (object-based attention, for a review *see* [55]). Given that so little is known about the mechanisms by which auditory objects are coded [56], we shall focus on those studies of auditory feature-based attention. A sound can be defined according to many different feature dimensions including frequency spectrum, temporal envelope, periodicity, spatial location, sound level, and duration. The experimenter can instruct listeners to attend to any feature dimension in order to investigate the effect of selective attention on the neural coding of that feature. Different listening conditions have been used for comparison with the “attend” condition. The least controlled of these is a passive listening condition in which participants are not given any explicit task instructions [30, 57, 58]. Even if there are cases where a task is required, but the cognitive demand of that task is low, participants are able to divide their attention across both relevant and irrelevant stimulus dimensions (*see* [59] for a review on attentional load). Again, this leads to an uncontrolled experimental situation. For greater control, some studies have employed a visual distractor task to compete for attentional resources and pull selective attention away from the auditory modality [60, 61]. However, there is some evidence that the mere presence of a visual stimulus exerts a significant influence on auditory cortical responses [62, 63] and hence modulation related to selective attention might interact with that related to the presence of visual stimuli in a rather complex manner. This can make comparison between the results from bimodal studies [60, 61] and unimodal auditory studies [32, 64] somewhat problematic.

One paradigm that has been commonly used to examine feature-based attention manipulates two different feature dimensions independently within the same experimental session and listeners are required to make a discrimination judgement to one feature or the other. Studies have compared attention to spatial features such as location, motion, and ear of presentation with attention to nonspatial features such as pitch and phonemes [60,

64]. Results typically demonstrate a response enhancement in nonprimary auditory regions. For example, Degerman et al. [60] found auditory enhancement in left posterior nonprimary regions, but only for attending to location relative to pitch and not the other way round. Ahveninen et al. [64] used a novel paradigm in which they measured the effect of attention on neural adaptation. Their fMRI results showed smaller adaptation effects in the right posterior nonprimary auditory cortex when attending to location (relative to phonemes), but again not the other way round. Both studies reported enhancement for attending to location in additional nonauditory regions, notably the prefrontal and right parietal areas. This asymmetry in the effects observed across spatial and nonspatial attended domains is worthy of further exploration since spatial analysis is well known to engage the right posterior auditory and right parietal cortex [65].

Another experimental design that has been used to examine feature-based attention presents concurrent visual and auditory stimuli and participants are required to make a discrimination judgement to stimuli in one modality or the other. One example of this design used novel melodies and geometric shapes, and participants were required to respond to either long note targets or vertical line targets [57, 58]. When “attending to the shapes” was subtracted from “attending to the melodies” the results revealed relative enhancement bilaterally in the lower boundary of the superior temporal gyrus. This finding supports the view that there is sensory enhancement when attending to the auditory modality. In addition, it was shown that when “attending to the shapes,” the auditory response was suppressed relative to a bimodal passive condition. This is tentative evidence for neural suppression when ignoring the auditory modality. A novel feature of the experiment by Degerman et al. [66] was that in one selective attention condition, participants had to respond to a target defined by a particular combination of cross-modal features (e.g. high pitch and red circle). The conventional general linear analysis did not show any significant difference in the magnitude of the auditory response in the cross-modal condition compared with a condition in which participants simply attended to the high- and low-pitch targets in the audiovisual stimulus. However, a region of interest analysis (defining a region in the posterolateral superior temporal gyrus) did suggest some enhancement for the audiovisual attention condition compared with the auditory attention condition. Thus, it is possible that nonprimary auditory regions are involved in attention-dependent binding of synchronous auditory and visual events into coherent audio–visual objects.

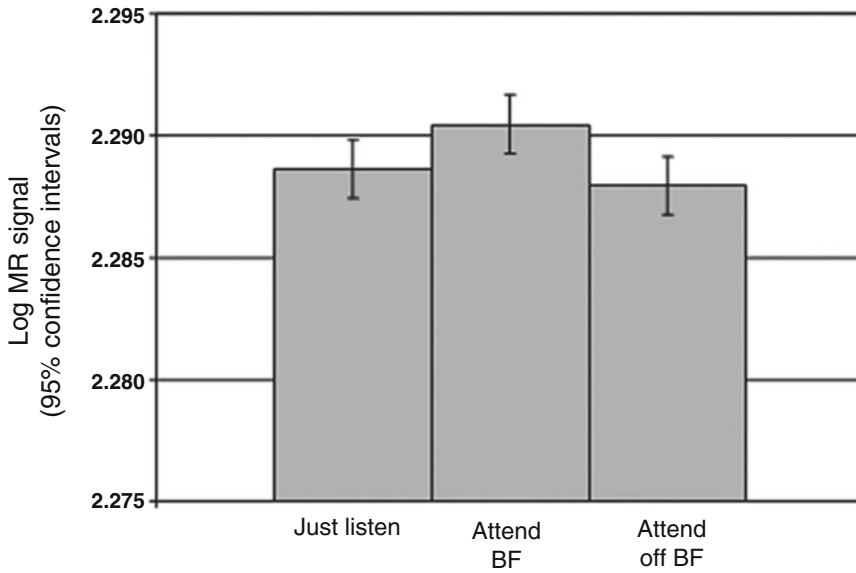
In audition, it has long been established behaviourally that when participants expect a tone at a specific frequency, their ability to detect a tone in a noise masker is significantly better when the tone is at the expected frequency than when it is at an unexpected frequency (the probe-signal paradigm [67]). The benefit of

selective attention for signal detection thresholds can be plotted as a function of frequency. The ability to detect tones at frequencies close to the expected frequency is also enhanced, and this benefit drops off smoothly with the distance away from the expected frequency [67, 68]. The width of this attention-based listening band is comparable to the width of the critical band related to the frequency-tuning curve, which can be measured psychophysically using notched noise maskers [68]. This equivalence suggests that selective attention might be operating at the level of the sensory representation of tone frequency.

Evidence from electrophysiological recordings demonstrates frequency-specific attentional modulation at the level of the primary auditory cortex, consistent with a neural correlate of the psychophysical phenomena found in the probe-signal paradigm. In a series of experiments, awake behaving ferrets were trained to perform a number of spectral tasks including tone detection and frequency discrimination [69]. In the tone detection task, ferrets were trained to identify the presence of a tone against a background of broadband rippled noise. The spectro-temporal receptive fields measured during the noise for frequency-tuned neurons showed strong facilitation around the target frequency that persisted for 30–40 ms. In the two-tone discrimination task, ferrets performed an oddball task in which they responded to an infrequent target frequency. Again, the spectro-temporal receptive fields showed an enhanced and persistent response for the target frequency, plus a *decreased* response for the reference frequency. These opposite effects serve to magnify the contrast between the two center frequencies, and thus facilitate the selection of the target. The results of these two tasks confirm that the acoustic filter properties of auditory cortical neurons can dynamically adapt to the attentional focus of the task.

Recently, we have addressed the question of attentional enhancement for selective attention to frequency using a high-resolution scanning protocol (1.5 mm<sup>2</sup> × 2.5 mm) (unpublished data). To control for the demands on selective attention, we presented two concurrent streams (low- and high-frequency tones). Participants were requested to attend to one frequency stream or the other and these attend conditions were presented in an interleaved manner throughout the experiment. Behavioral testing confirmed that performance significantly deteriorated when these sounds were presented in a divided attention task. To be able to identify high- and low-frequency sensitive areas around the primary auditory cortex we designed two types of stimuli using different rhythms for each of the two streams. For example, one stimulus contained a “fast” high-frequency rhythm and a “slow” low-frequency rhythm so that the stimulus contained a majority of high-frequency tones. The other stimulus was the converse. Areas of high-frequency sensitivity were identified by subtracting the low-frequency majority stimulus from the high-frequency majority stimulus, and vice-versa (Fig. 9c, d). For each of the three





**Fig. 10** Response to the three listening conditions: just listen, attend to the best frequency (BF) tones, and attend OFF BF. The data shown are for those stimuli in which BF tones formed the majority (80%) of the tones in the sound sequence, combining responses across areas 1–3. The error bars denote the 95% confidence intervals

participants, we selected those frequency-specific areas that best corresponded to areas 1–4 (defined by Talavage et al. [51, 52]; see Sect. 2.1). Within these areas, we extracted the BOLD signal time course for every voxel and performed a log transform to standardize the data. We collapsed the data across low- and high-frequency sensitive areas [1–4] according to their “best frequency” (BF). The best frequency of an area corresponds to the frequency that evokes the largest BOLD response. A univariate ANOVA showed response enhancement when participants were attending to the BF of that area, compared with attending to the other frequency ( $p < 0.01$ ). In addition, response enhancement was also found when attending to the BF of that area, compared with passive listening ( $p < 0.05$ ) (Fig. 10). Note that for these results area 4 was excluded from the analysis, because it showed different pattern of attentional modulation. The response profile of area 4 might differ from that of areas 1–3 in other ways because it is not consistently present in all listeners [51]. Our finding of frequency-specific attentional enhancement in primary auditory regions contrasts with that of Petkov et al. [61], who reported attention-related modulation to be independent of stimulus frequency and to engage mainly the nonprimary auditory areas. However, our result is more in keeping with the predictions made by the neurophysiological data reported by Fritz et al. [69].

---

### 3 Future Directions

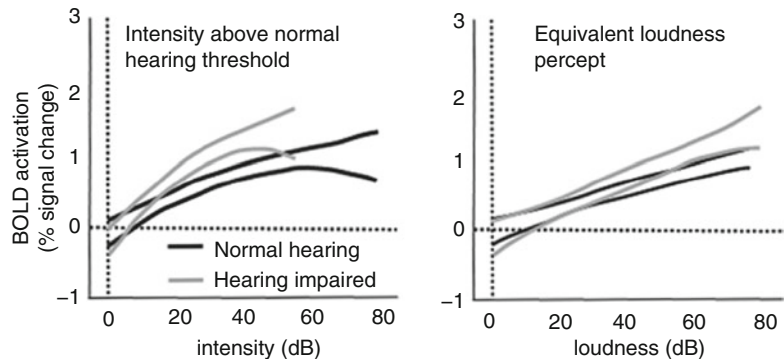
It is increasingly likely that auditory cortical regions compute aspects of the sound signal that are more complicated in their nature than the simple physical acoustic attributes of the sound. Thus, the encoded features of the sound reflect an increasingly abstract representation of the sound stimulus. We have already presented some evidence for this in terms of the way in which the auditory cortical response is exquisitely sensitive to the temporal context of the sound, particularly the way in which the time course of the BOLD response represents the temporal envelope characteristics of the sound, including sound onsets and offsets [18, 40], (*see* Sect. 1.4). However, there are many other ways in which neural coding reflects higher level processing. In this final section, we shall introduce two important aspects of the listening context that determine the auditory BOLD response: the perceptual experience of the listener and the operational aspects of the task. A number of fMRI studies have demonstrated ways in which activity within the human auditory cortex is modulated by auditory sensations, including loudness, pitch, and spatial width. Other studies have revealed that task relatedness is also a significant determining factor for the pattern of activation. These findings highlight how future auditory fMRI studies could usefully investigate these contributory factors in order to provide a more complete picture of the neural basis of the listening process.

#### **3.1 Cortical Activation Reflects Perceptually Relevant Coding**

One approach used in auditory fMRI to investigate perceptually relevant coding imposes systematic changes to the listener's perception of a sound signal by parametrically manipulating certain acoustic parameters and subsequently correlating the perceptual change with the variation in the pattern of activation. For example, by increasing sound intensity (measured in SPL), one also increases its perceived loudness (measured in phons). Loudness is a perceptual phenomenon that is a function of the auditory excitation pattern induced by the sound, integrated across frequency. Sound intensity and loudness are measures of different phenomena. For example, if the bandwidth of a broadband signal is increased while its intensity is held constant, then loudness nevertheless increases because the signal spans a greater number of frequency channels. In an early fMRI study, Hall et al. [31] presented single-frequency tones and harmonic-complex tones that were matched either in intensity or in loudness. The results showed that the complex tones produced greater activation than did the single-frequency tones, irrespective of the matching scheme. This result indicates that bandwidth had a greater effect on the pattern of auditory activation than sound level. Nevertheless, when the data were collapsed across stimulus class, the amount of activation was significantly correlated with the loudness scale, not with the intensity scale.

In people with elevated hearing thresholds, the perception of sound level is distorted. They typically experience the same dynamic range of loudness as normally hearing listeners despite having a compressed range of sensitivity to sound level. The BOLD response to sound level is reflected in a disproportionate increase in loudness with intensity. A recent study has characterized the BOLD response to frequency-modulated tones presented at a broad range of intensities (0–70 dB above the normal hearing threshold) [33]. Both normally hearing and hearing impaired groups showed the same steepness in the linear increase in auditory activation as a function of loudness, but not of intensity (Fig. 11). The results from this study clearly demonstrate that the BOLD response can be interpreted as a correlate of the subjective strength of the stimulus percept.

Pitch can be defined as the sensation whose variation is associated with musical melodies. Together with loudness, timbre, and spatial location, pitch is one of the primary auditory sensations. The salience of a pitch is determined by several physical properties of the pitch signal, one being the numbered harmonic components comprising a harmonic-complex tone. The cochlea separates out the frequency components of sounds to a limited extent, so that the first eight harmonics of a harmonic-complex tone excite distinct places in the cochlea and are said to be “resolved,” whereas the higher harmonics are not separated and are said to be “unresolved.” Pitch discrimination thresholds for unresolved harmonics are substantially higher than those for resolved harmonics, consistent with the former type of stimulus evoking a less salient pitch



**Fig. 11** Growth in the blood oxygen level-dependent (BOLD) response (measured in % signal change) for high (4–8 kHz) frequency-modulated tone presented across a range of sound levels for ten normally hearing participants (*black lines*) and ten participants with a high-frequency hearing loss (*gray lines*). The *left hand panel* shows the growth as a function of sound intensity, while the *right hand panel* plots the same data as a function of the equivalent loudness percept. The *upper and lower lines* denote the 95% confidence interval of the quadratic polynomial fit to the data. This graph summarizes data presented by Langers et al [33]

[70]. A pairwise comparison between the activation patterns for resolved (strong pitch) and unresolved (weak pitch) harmonic-complex tones has identified differential activation in a small, spatially localized region of nonprimary auditory cortex, overlapping the anterolateral end of Heschl's gyrus [71]. The authors claim that this finding reflects the cortical representation for pitch salience. Another way to determine the salience of a pitch is by the degree of fine temporal regularity in the stimulus (i.e. the monaural repeating pattern within frequency channels). This is true even for signals in which there are no distinct frequency peaks in the cochlear excitation pattern from which to calculate the pitch. A range of pitch saliencies can be created by parametrically varying the degree of temporal regularity in an iterated-ripple noise stimulus (using 0, 1, 2, 4, 8, and 16 add-and-delay iterations during stimulus generation) [72]. Again the anterolateral end of Heschl's gyrus appeared highly responsive to the change in pitch salience, in a linear manner.

Spatial location is another important auditory sensation that is determined by the fine temporal structure in the signal, this time it being the binaural temporal characteristics across the two ears. The interaural correlation (IAC) of a sound represents the similarity between the signals at the left and right ears. Changes in the IAC of a wideband signal result in changes in sound's perceived "width" when presented through headphones. A noise with an IAC of 1.0 is typically perceived as sound with a compact source located at the center of the head. As the IAC is reduced the source broadens. For an IAC of 0.0, it eventually splits into two separate sources, one at each ear [73]. Again the parametric approach has been employed to measure activation across a range of IAC values (1.00, 0.93, 0.80, 0.60, 0.33, and 0.00) [74]. The authors found a significant positive relationship between BOLD activity and IAC, which was confined to the anterolateral end of Heschl's gyrus, the region that is also responsive to pitch salience. The slope of the function was not precisely linear but the BOLD response was more sensitive to changes in IAC at values near to unity than at values near zero. This response pattern is qualitatively compatible with previous behavioral measures of sensitivity to IAC [75].

There is some evidence to support the claim that the neural representations of auditory sensations (including loudness, pitch, and spatial width) evolve as one ascends the auditory pathway. Budd et al. [74] examined sensitivity to values of IAC associated with spatial width within the inferior colliculus, the medial geniculate nucleus, as well as across different auditory cortical regions, but the effects were significant only within the nonprimary auditory cortex. Griffiths et al. [72, 76] also examined sensitivity to the increases in temporal regularity associated with pitch salience within the cochlear nucleus, inferior colliculus, medial geniculate nucleus, as well as across different auditory cortical regions. Some

degree of sensitivity to pitch salience was found at all sites, but the preference appeared greater in the higher centers than in the cochlear nucleus [76]. Thus, the evidence supports the notion of an increasing responsiveness to percept attributes of sound throughout the ascending auditory system, culminating in the nonprimary auditory cortex. These findings are consistent with the hierarchical processing of sound attributes.

Encoding the perceptual properties of a sound is integral to identifying the object properties of that sound source. The nonprimary auditory cortex probably plays a key role in this process because it has widespread cortical projections to frontal and parietal brain regions and is therefore ideally suited to access distinct higher level cortical mechanisms for sound identification and localization. Recent trends in auditory neuroscience are increasingly concerned with auditory coding beyond the conventional limits of the auditory cortex (the superior temporal gyrus in humans), particularly with respect to the hierarchical organization of sensory coding via dorsal and ventral auditory processing routes. At the top of this hierarchy stands the brain's representation of an auditory "object." The concept of an auditory object still remains controversial [56]. Although it is clear that the brain needs to code information about the invariant properties of a sound source, research in this field is considerably underdeveloped. Future directions are likely to begin to address critical issues such as the definition of an auditory object, whether the concept is informative for auditory perception, and optimal paradigms for studying object coding.

### **3.2 Cortical Activation Also Reflects Behaviourally Relevant Coding**

Listeners interact with complex auditory environments that, at any one time point, contain multiple auditory objects located at dynamically varying spatial locations. One of the primary challenges for the auditory system is to analyze this external environment in order to inform goal-directed behavior. In Sect. 2.2 we introduced some of the neurophysiological evidence for the importance of the attentional focus of the task in determining the pattern of auditory cortical activity [69]. Here, we consider the contribution of human auditory fMRI research to this question. In particular, we present the interesting findings of one group who have started to address how the auditory cortex responds to the context and the procedural and cognitive demands of the listening task (*see* [77] for a review).

In that review, Scheich and colleagues report a series of research studies in which they suggest that the function of different auditory cortical areas is not determined so much by stimulus features (such as timbre, pitch, motion, etc.), but rather by the task that is performed. For example, one study reported the results of two fMRI experiments in which the same frequency-modulated stimuli were presented under different task conditions [78]. Top-down influences strongly affected the strength of the auditory response.

When a pitch-direction categorization task was compared with passive listening, a greater response was found in right posterior nonprimary auditory areas (planum temporale). Moreover, hemispheric differences were also found when comparing the response to two different tasks. The right nonprimary areas responded more strongly when the task required a judgement about pitch direction (rising or falling), whereas the left nonprimary areas responded more strongly when the task required a judgement about the sound duration. It is not the case that the right posterior nonprimary areas were *only* engaged by sound categorization because this region was more responsive to the critical sound feature (frequency modulation) during passive listening than were surrounding auditory areas (*see* also ref. [32]). These results broadly indicate an interaction between the stimulus and the task, which influences the pattern of auditory cortical activity. The precise characteristics of this interaction are worthy of future studies.

## References

- Palmer AR, Bullock DC, Chambers JD (1998) A high-output, high-quality sound system for use in auditory fMRI. *Neuroimage* 7:S357
- Chou CK, McDougall JA, Chan KW (1995) Absence of radiofrequency heating from auditory implants during magnetic-resonance imaging. *Bioelectromagnetics* 16(5):307–316
- Heller JW, Brackmann DE, Tucci DL, Nyenhuis JA, Chou CK (1996) Evaluation of MRI compatibility of the modified nucleus multichannel auditory brainstem and cochlear implants. *Am J Otol* 17(5):724–729
- Shellock FG, Morisoli S, Kanal E (1993) MR procedures and biomedical implants, materials, and devices—1993 update. *Radiology* 189(2):587–599
- Weber BP, Neuburger J, Battmer RD, Lenarz T (1997) Magnetless cochlear implant: relevance of adult experience for children. *Am J Otol* 18(6):S50–S51
- Wild DC, Head K, Hall DA (2006) Safe magnetic resonance scanning of patients with metallic middle ear implants. *Clin Otolaryngol* 31(6):508–510
- Giraud AL, Truy E, Frackowiak R (2001) Imaging plasticity in cochlear implant patients. *Audiol Neurootol* 6(6):381–393
- Moelker A, Wielopolski PA, Pattynama PM (2003) Relationship between magnetic field strength and magnetic-resonance-related acoustic noise levels. *MAGMA* 16:52–55
- Foster JR, Hall DA, Summerfield AQ, Palmer AR, Bowtell RW (2000) Sound-level measurements and calculations of safe noise dosage during fMRI at 3T. *J Magn Reson Imaging* 12:157–163
- Ravicz ME, Melcher JR (2001) Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am* 109(1):216–231
- Price DL, De Wilde JP, Papadaki AM, Curran JS, Kitney RI (2001) Investigation of acoustic noise on 15 MRI scanners from 0.2 T to 3 T. *J Magn Reson Imaging* 13(2):288–293
- Hedeen RA, Edelstein WA (1997) Characterization and prediction of gradient acoustic noise in MR imagers. *Magn Reson Med* 37(1):7–10
- Hennel F, Girard F, Loenneker T (1999) “Silent” MRI with soft gradient pulses. *Magn Reson Med* 42:6–10
- Brechmann A, Baumgart F, Scheich H (2002) Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. *J Neurophysiol* 87:423–433
- Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26(2):212–215
- Assmann P, Summerfield Q (2004) Perception of speech under adverse conditions. In: Greenberg S, Ainsworth WA, Popper AN, Fay RR (eds) *Speech processing in the auditory system*. Springer, New York, pp 231–308
- Healy EW, Moser DC, Morrow-Odom KL, Hall DA, Fridriksson J (2007) Speech perception in MRI scanner noise by persons with aphasia. *J Speech Lang Hear Res* 50:323–334
- Harms MP, Melcher JR (2002) Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 88:1433–1450

19. Chambers JD, Akeroyd MA, Summerfield AQ, Palmer AR (2001) Active control of the volume acquisition noise in functional magnetic resonance imaging: method and psychoacoustical evaluation. *J Acoust Soc Am* 110(6):3041–3054
20. Hall DA, Chambers J, Foster J, Akeroyd MA, Coxon R, Palmer AR (2009) Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *J Acoust Soc Am* 125(1):347–359
21. Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) ‘Sparse’ temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223
22. Bandettini PA, Jesmanowicz A, Van Kylen J, Birn RM, Hyde JS (1998) Functional MRI of brain activation induced by scanner acoustic noise. *Magn Reson Med* 39:410–416
23. Bilecen D, Scheffler K, Schmid N, Tschopp K, Seelig J (1998) Tonotopic organization of the human auditory cortex as detected by BOLD-fMRI. *Hear Res* 126:19–27
24. Hall DA, Summerfield AQ, Gonçalves MS, Foster JR, Palmer AR, Bowtell RW (2000) Time-course of the auditory BOLD response to scanner noise. *Magn Reson Med* 43:601–606
25. Shah NJ, Jäncke L, Grosse-Ruyken M-L, Müller-Gärtner HW (1999) Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. *J Magn Reson Imaging* 9(1):19–25
26. Talavage TM, Edmister WB, Ledden PJ, Weisskoff RM (1999) Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Hum Brain Mapp* 7(2):79–88
27. Elliott MR, Bowtell RW, Morris PG (1999) The effect of scanner sound in visual, motor, and auditory functional MRI. *Magn Reson Med* 41(6):1230–1235
28. Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 7:89–97
29. Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40(4):859–869
30. Hall DA, Haggard MP, Akeroyd MA, Summerfield AQ, Palmer AR, Elliott MR, Bowtell RW (2000) Modulation and task effects in auditory processing measured using fMRI. *Hum Brain Mapp* 10(3):107–119
31. Hall DA, Haggard MP, Summerfield AQ, Akeroyd MA, Palmer AR, Bowtell RW (2001) Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. *J Acoust Soc Am* 109(4):1559–1570
32. Hart HC, Palmer AR, Hall DA (2004) Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Hum Brain Mapp* 21:178–190
33. Langers DRM, Backes WH, Van Dijk P (2007) Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage* 34:264–273
34. Schwarzbauer C, Davis MH, Rodd JM, Johnsrude I (2006) Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. *Neuroimage* 29(3):774–782
35. Bregman AS (1990) Auditory scene analysis: the perceptual organisation of sound. MIT, Cambridge, MA
36. Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* 116(3):1656–1670
37. Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 151:167–187
38. Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 77:923–943
39. Bartlett EL, Wang X (2005) Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 94:83–104
40. Harms MP, Guinan JJ, Sigalovsky IS, Melcher JR (2005) Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. *J Neurophysiol* 93:210–222
41. Palmer AR, Hall DA, Sumner C, Barrett DJK, Jones S, Nakamoto K, Moore DR (2007) Some investigations into non-passive listening. *Hear Res* 229:148–157
42. Hall DA, Edmondson-Jones M, Fridriksson J (2006) Periodicity and frequency coding in human auditory cortex. *Eur J Neurosci* 24:3601–3610
43. Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140–149
44. Hart HC, Hall DA, Palmer AR (2003) The sound-level-dependent growth in the extent of fMRI activation in Heschl’s gyrus is different for low- and high-frequency tones. *Hear Res* 179(1–2):104–112
45. Von Békésy G (1947) The variations of phase along the basilar membrane with sinusoidal vibrations. *J Acoust Soc Am* 19:452–460

46. Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386:304–316
47. Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50:275–296
48. Petkov CL, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4(7):213–226
49. Hall DA, Hart HC, Johnsrude IS (2003) Relationships between human auditory cortical structure and function. *Audiol Neurootol* 8(1):1–18
50. Schönwiesner M, Von Cramon DY, Rubsamen R (2002) Is it tonotopy after all? *Neuroimage* 17:1144–1161
51. Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR (2000) Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res* 150:225–244
52. Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91:1282–1296
53. Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci* 23(1):315–341
54. Maunsell JHR, Treue S (2006) Feature-based attention in visual cortex. *Trends Neurosci* 29(6):317–322
55. Scholl BJ (2001) Objects and attention: the state of the art. *Cognition* 80(1–2):1–46
56. Griffiths TD, Warren JD (2004) What is an auditory object? *Nat Rev Neurosci* 5(11):887–892
57. Johnson JA, Zatorre RJ (2005) Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cereb Cortex* 15(10):1609–1620
58. Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *Neuroimage* 31(4):1673–1681
59. Lavie N (2005) Distracted and confused? Selective attention under load. *Trends Cogn Sci* 9(2):75–82
60. Degerman A, Rinne T, Salmi J, Salonen O, Alho K (2006) Selective attention to sound location or pitch studied with fMRI. *Brain Res* 1077(1):123–134
61. Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. *Nat Neurosci* 7(6):658–663
62. Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27(8):1824–1835
63. Lehmann C, Herdener M, Esposito F, Hubl D, di Salle F, Scheffler K, Bach DR, Federspiel A, Kretz R, Dierks T, Seifritz E (2006) Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *Neuroimage* 31(1):294–300
64. Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin F-H, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci U S A* 103(39):14608–14613
65. Lewald J, Meister IG, Weidemann J, Topper R (2004) Involvement of the superior temporal cortex and the occipital cortex in spatial hearing: evidence from repetitive transcranial magnetic stimulation. *J Cogn Neurosci* 16(5):828–838
66. Degerman A, Rinne T, Pekkola J, Autti T, Jaaskelainen IP, Sams M, Alho K (2007) Human brain activity associated with audiovisual perception and attention. *Neuroimage* 34(4):1683–1691
67. Greenberg GZ, Larkin WD (1968) Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J Acoust Soc Am* 44(6):1513–1523
68. Schlauch RS, Hafer ER (1991) Listening bandwidths and frequency uncertainty in pure-tone signal detection. *J Acoust Soc Am* 90(3):1332–1339
69. Fritz JB, Elhilali M, David SV, Shamma SA (2007) Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in AI? *Hear Res* 229:186–203
70. Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95:3529–3540
71. Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24(30):6810–6815
72. Griffiths TD, Büchel C, Frackowiak RSJ, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. *Nat Neurosci* 1:422–427
73. Blauert J, Lindemann W (1986) Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J Acoust Soc Am* 79(3):806–813
74. Budd TW, Hall DA, Goncalves MS, Akeroyd MA, Foster JR, Palmer AR, Head K, Summerfield AQ (2003) Binaural



- specialisation in human auditory cortex: an fMRI investigation of interaural correlation sensitivity. *Neuroimage* 20(3):1783–1794
75. Culling JF, Colburn HS, Spurchise M (2001) Interaural correlation sensitivity. *J Acoust Soc Am* 110(2):1020–1029
76. Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci* 4:633–637
77. Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW (2007) The cognitive auditory cortex: task-specificity of stimulus representations. *Hear Res* 229:213–224
78. Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb Cortex* 15(5):578–587