3-12-2008

Functional imaging of the auditory processing applied to speech sounds

Roy D Patterson

Ingrid Johnsrude

PHILOSOPHICAL
TRANSACTIONS
OF
THE ROYAL
SOCIETY B

Phil. Trans. R. Soc. B (2008) **363**, 1023–1035
doi:10.1098/rstb.2007.2157
Published online 7 September 2007

Functional imaging of the auditory processing applied to speech sounds

Roy D. Patterson[1,*] **and Ingrid S. Johnsrude**[2]

[1]*Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK*
[2]*Department of Psychology, Queen's University, 62 Arch Street, Kingston, Ontario, Canada K7L 3N6*

In this paper, we describe domain-general auditory processes that we believe are prerequisite to the linguistic analysis of speech. We discuss biological evidence for these processes and how they might relate to processes that are specific to human speech and language. We begin with a brief review of (i) the anatomy of the auditory system and (ii) the essential properties of speech sounds. Section 4 describes the general auditory mechanisms that we believe are applied to all communication sounds, and how functional neuroimaging is being used to map the brain networks associated with domain-general auditory processing. Section 5 discusses recent neuroimaging studies that explore where such general processes give way to those that are specific to human speech and language.

Keywords: auditory anatomy; speech sounds; auditory processing; neuroimaging of pitch; neuroimaging of speech sounds

1. INTRODUCTION

Speech is a rich social signal that conveys a wealth of information. Not only is it a linguistic signal, used to communicate information and ideas, but it also contains non-linguistic information about the size, sex, background, social status and emotional state of the speaker. Finally, it is usually experienced as a multisensory and interactive signal; these are important aspects that also do not fall within the traditional realm of linguistic analysis. These non-linguistic aspects of communication are a reminder that speech shares characteristics with communication in other animals, including other primates. The initial stages of auditory processing, which rely on a neural organization that is evolutionarily conserved among many primate species, are probably general and apply to all communication sounds, not just to speech. Accordingly, we begin with a brief overview of primate anatomy. At the same time, the complexity of human communication indicates that it engages additional neural apparatus subserving linguistic and social cognition. The point in the system where the processing radiates out into divergent functions is the topic of §5.

2. A BRIEF OVERVIEW OF AUDITORY ANATOMY

(a) *The subcortical auditory system in humans*

In humans, the principal components of the subcortical auditory system lie in a frontal plane that extends from the ear canal to the upper surface of the central portion of the temporal lobe. Between the cochlea and the auditory cortex, there are four major centres of neural processing: the cochlear nucleus (CN); the superior olivary complex (SOC); the inferior colliculus (IC); and the medial geniculate body (MGB) of the thalamus. Work in other primates suggests that there are mandatory synapses for auditory processing in three of the four nuclei (CN, IC and MGB), which supports the view that these nuclei perform transformations that are applied to all sounds as they proceed up the pathway, much as the cochlea performs a mandatory frequency analysis on all sounds entering the auditory system. In the visual system, there is only one synapse between the retina and visual cortex in the lateral geniculate nucleus.

Information from the two ears is probably integrated in several nuclei in the subcortical auditory system. The CN projects to both the contralateral and the ipsilateral SOC, where minute differences in the timing of the versions of a sound at the two ears are correlated, permitting estimation of source location. The CN also projects to both contralateral and ipsilateral IC, and the two ICs are themselves densely interconnected. Thus, the subcortical auditory system does not maintain a clear segregation of information by the ear of entry. In contrast, in the visual system, there is no binocular processing prior to visual cortex. The complexity of the subcortical auditory system is probably due, at least in part, to the temporal precision of the neural representation of sound (Patterson *et al.* 1999). Auditory nerve fibres between the cochlea and the CN fire in phase with basilar membrane motion up to approximately 5000 Hz, and the nuclei that process this sub-millisecond information must be close to the source to minimize temporal distortion. The maximum rate of phase locking drops to approximately 500 Hz in the IC, and to approximately 50 Hz in the MGB and primary auditory cortex (PAC), which suggests that the form of the neural code changes at least twice as the information progresses from cochlea to cortex, once at the level of the IC and once at the level of the MGB.

[*] Author for correspondence (rdp1@cam.ac.uk).

One contribution of 13 to a Theme Issue 'The perception of speech: from sound to meaning'.

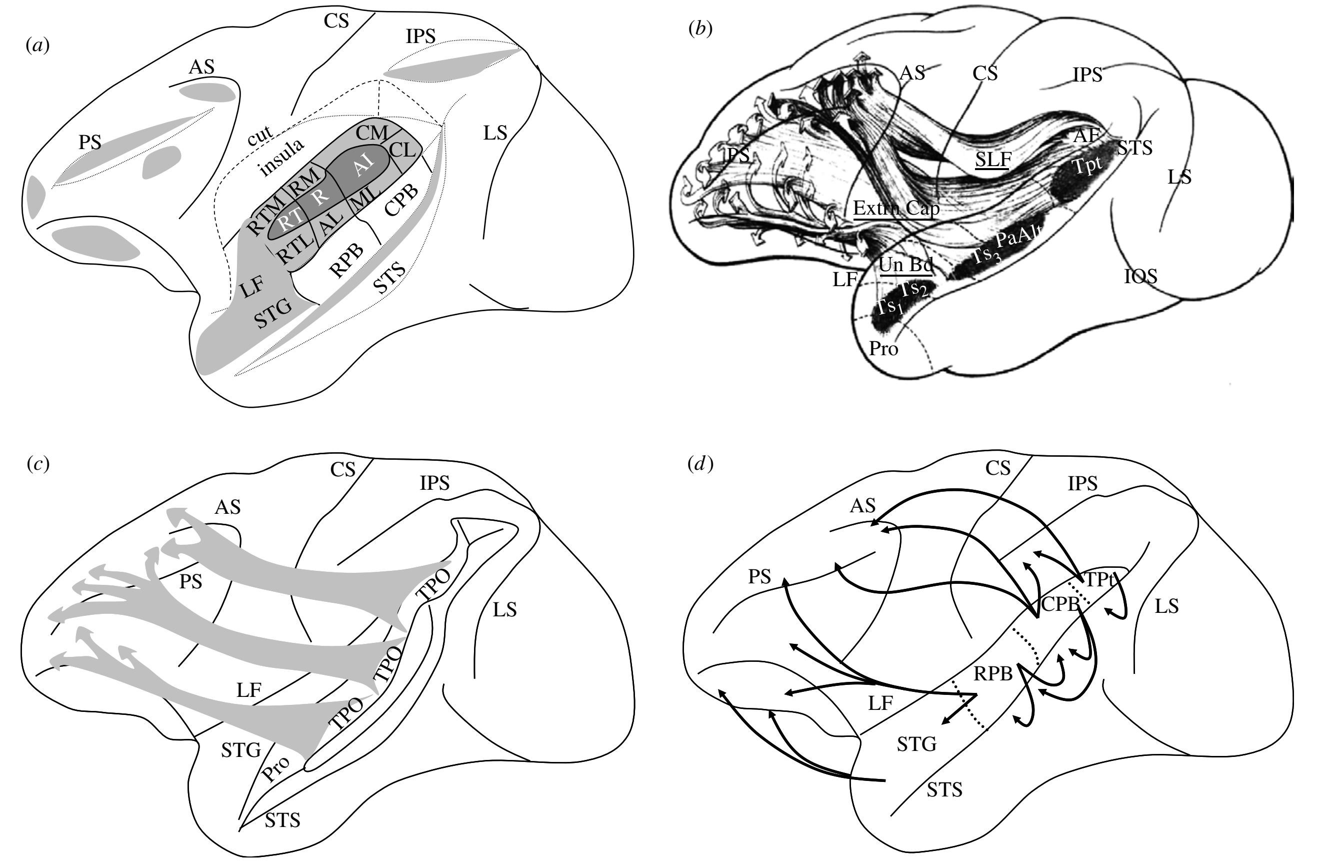

Figure 1. Four representations of the anatomical connections of the temporal lobe in the macaque brain. (*a*) The anatomical organization of the auditory cortex is consistent with at least four levels of processing, including core regions (darkest shading) belt regions (lighter shading), parabelt regions (stripes) and temporal and frontal regions that interconnect with belt and parabelt (lighter shading). (Adapted from Kaas *et al.* (1999) and Hackett & Kaas (2004)). Dotted lines indicate sulci that have been opened to show auditory regions. Regions along the length of (*b*) superior temporal gyrus and (*c*) dorsal bank of the superior temporal sulcus connect with prefrontal regions in a topographically organized anterior-to-posterior fashion. (*b*) Adapted from Petrides & Pandya (1988, p. 64); (*c*) adapted from Seltzer & Pandya (1989*a*). (*d*) Connectivity of auditory belt and parabelt; adapted from Hackett & Kaas (2004). AF, arcuate fasciculus; AS, arcuate sulcus; CS, central sulcus; Extm Cap, extreme capsule; IOS, inferior occipital sulcus; IPS, intraparietal sulcus; LF, lateral fissure; LS, lunate sulcus; PS, principal sulcus; SLF, superior longitudinal fasciculus; STG, superior temporal gyrus; STS, superior temporal sulcus; UnBd, uncinate bundle. (*Note*. Abbreviations are not spelt out if they are the conventional label for a microanatomically or physiologically defined area).

(b) *The anatomy of auditory cortex and its projections, in the macaque*

In humans, the principal components of the cortical auditory system are not well understood. Microelectrode recordings, the cornerstone of non-human neurophysiology, can only be undertaken in rare circumstances (e.g. during neurosurgery; Howard *et al.* 2000; Brugge *et al.* 2003). Post-mortem histological material is scarce and of relatively poor quality (Hackett *et al.* 2001; Wallace *et al.* 2002), and *in vivo* tracer studies in humans are currently not possible. The rhesus macaque monkey (*Macaca mulatta*) provides an animal model for the organization of auditory cortex (Rauschecker *et al.* 1997; Rauschecker 1998; Kaas *et al.* 1999; Kaas & Hackett 2000), and this can be supplemented by the (relatively few) anatomical and neurophysiological studies that have been conducted in humans (Liegeois-Chauvel *et al.* 1991; Rivier & Clarke 1997; Howard *et al.* 2000; Hackett *et al.* 2001; Morosan *et al.* 2001; Rademacher *et al.* 2001; Wallace *et al.* 2002; see Hall *et al.* (2003) and Scott & Johnsrude (2003), for reviews).

A note of caution must be sounded in assuming anatomical and functional homologies between macaques and humans. Most obviously, functional specialization must diverge in the two species at, or before, the point where speech-specific processing begins in humans. Furthermore, unlike our own species, vocalization is not an important form of communication in macaques. Also, auditory research in the macaque has been largely restricted to experiments with very simple sounds such as clicks and pure tones, which may not require extensive cortical processing. As a result, the functional specialization of the core, belt and parabelt regions is simply not known, and macaque research provides only the most general indication of where to look for specific forms of processing in humans.

The organization in the macaque is shown in figure 1*a*. Cortical afferents from the ventral division of the MGB project to three tonotopically organized fields on the superior temporal gyrus (STG; Rauschecker *et al.* 1997; Kaas *et al.* 1999; Kaas & Hackett 2000). This 'core' of primary areas projects to a surrounding 'belt' of anatomically distinguishable cortical fields which exhibit interconnections among adjacent regions (Merzenich & Brugge 1973; Pandya & Sanides 1973; Jones *et al.* 1995; Pandya 1995; Hackett *et al.* 1998; Rauschecker 1998; Kaas & Hackett 2000;

Rauschecker & Tian 2000). Belt areas connect with lateral 'parabelt' fields, again through connections between physically adjacent regions. The hierarchical connections of the core, belt and parabelt areas suggest at least three discrete levels of processing in the macaque (Pandya 1995; Hackett *et al.* 1998; Rauschecker 1998; Kaas *et al.* 1999; Kaas & Hackett 2000; Rauschecker & Tian 2000).

Recent neuroimaging studies (reviewed in §5) indicate that the superior temporal sulcus (STS) region in humans is important for speech-sound perception. Drawing inferences from macaque cortical organization is problematic in the STS, since humans have a middle temporal gyrus (including the ventral bank of the STS) and macaques do not. Human homologies of the ventral bank regions that have been mapped in the macaque are particularly uncertain. Nevertheless, the anatomical organization of the upper bank of the macaque STS may be somewhat conserved in humans, and it is currently the best evidence we have as to what to expect in human STS.

The STS in the macaque is anatomically heterogeneous, but much of its upper bank, running the length of the STS, comprises a region (area TAa) that receives its input mainly from auditory cortex (Seltzer & Pandya 1978, 1989*b*). This region projects into adjacent polysensory cortex in the depth of the STS, as well as to the inferior parietal lobule and prefrontal cortex (Seltzer & Pandya 1989*a*,*b*). Furthermore, anterior STS regions project to ventral and anterior frontal regions, and more posterior STS regions project to more posterior and dorsal frontal regions (and to parietal cortex; figure 1*b*). Similarly, as shown in figures 1*c*,*d* anterior belt and parabelt also interconnect directly, and in a topographically organized way, with multiple sites within orbitofrontal, ventrolateral and dorsolateral frontal cortex including Brodmann areas 46, 12 and 45 (Petrides & Pandya 1984; Hackett *et al.* 1998, 1999; Romanski *et al.* 1999*a*,*b*). Importantly, area 45 in humans, located in the inferior frontal gyrus (IFG; pars triangularis), is considered as one of the architectonic constituents of Broca's area (Amunts *et al.* 1999). This distributed set of fields in STG, STS, parietal and prefrontal cortex constitutes a potential fourth stage of processing (Kaas *et al.* 1999; figure 1*a*).

(c) *Links between perception and production in humans*

At the level of cortex, anatomical connectivity suggests that auditory perception and vocal production may be quite intimately linked. Auditory core, belt and parabelt regions all project into the dorsal caudate and putamen—components of the basal ganglia—which are traditionally considered to serve a primarily motor function (Yeterian & Pandya 1998). STS regions that receive projections from auditory cortices, in turn project to regions of the inferior parietal lobule that interconnect with motor cortex via premotor cortex (Pandya & Seltzer 1982; Seltzer & Pandya 1991; Petrides & Pandya 2002). Finally, Brodmann areas 45 and 46 in frontal cortex, which receive auditory projections, interconnect with motor regions via area 44 and premotor cortex.

Physiological data are consistent with a link between auditory perception and vocal production, and they indicate that the coupling is quite rapid. Matt Howard, John Brugge and colleagues have used depth electrode stimulation and electrophysiological recording in neurosurgical patients to explore the evoked responses and connectivity in a circuit involving PAC, a posterolateral region of the STG which they call posterior lateral superior temporal (PLST), IFG (pars triangularis and opercularis) and orofacial motor cortex (Garell *et al.* 1998; Howard *et al.* 2000; Brugge *et al.* 2003; Greenlee *et al.* 2004). Evoked responses in PAC of Heschl's gyrus (HG) had response latencies ranging from 15 to 25 ms, which are compatible with the magnetoencephalography (MEG) data on click latency in PAC reported by Lütkenhöner *et al.* (2003). Then, when this region of HG was electrically stimulated, it resulted in an evoked potential in PLST (Howard *et al.* 2000; Brugge *et al.* 2003). The average onset latency for this evoked response was only 2.0 ms, consistent with an ipsilateral corticocortical connection between HG and PLST. PLST appears to make a functional connection with the IFG (Garell *et al.* 1998) with onset latencies of approximately 10 ms, and cortical stimulation of posterior IFG elicits responses in orofacial motor cortex with onset latencies of approximately 6.0 ms (Greenlee *et al.* 2004). Taken together, these results suggest that a sound in the environment could, in principle, have an impact on neural activity in orofacial motor cortex within 35 ms of stimulus onset, and most of that time is spent in the pathway from the cochlea to PAC.

In summary, this overview of the anatomy of auditory cortex suggests that, following the succession of nuclei in the subcortical pathway, the information in auditory cortex radiates out in parallel paths from core areas, and cascades into at least three spatially distributed sets of regions, comprising at least three further processing stages. Other sense information is integrated with auditory information early on in cortical processing, and prominent feedback routes connect adjacent regions at all levels. Perceptual processes must depend on this anatomical organization.

Now we turn to the characteristics of speech sounds and describe a model of the processes that we believe are applied to all communication sounds before speech-specific processing begins in cortex.

3. GENERAL AUDITORY PROCESSES INVOLVED IN SPEECH PERCEPTION

When a child and an adult utter the 'same' syllable, it is only the linguistic message of the syllable that is the same. The child has a shorter vocal tract and lighter vocal cords, and as a result, the waveforms carrying the message are quite different for the child and the adult. Although humans have no difficulty in understanding that a child and an adult have said the same word, evaluating the equivalence is far from trivial, as indicated by the fact that speech-recognition machines find this task difficult. Indeed, when trained on the speech of a man, recognition machines are notoriously bad at understanding the speech of a woman, let alone a child. The robustness of auditory perception has led Irino & Patterson (2002) to

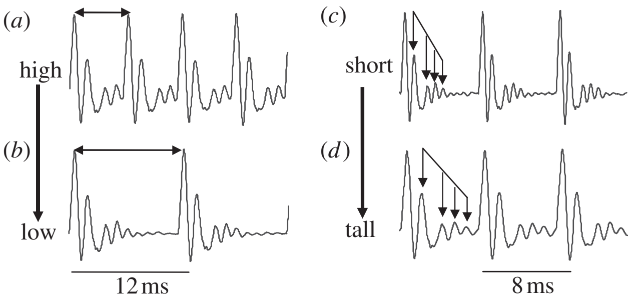

Figure 2. Internal structure of voiced sounds illustrating the size factors: pulse rate and resonance rate. (*a*,*b*) Glottal pulse rate and (*c*,*d*) vocal-tract length have a major effect on both the waveform and the spectrum of the sound, but human perception is extremely robust to changes in both of these factors.

hypothesize that the auditory system possesses mechanisms that automatically assess the vocal-tract length (VTL) and glottal pulse rate (GPR) of the speaker. Moreover, since humans produce speech sounds in much the same way as all other mammals, it is assumed that such mechanisms are part of the processing applied to all sounds. The value of this analysis is that it helps to produce a size-invariant representation of the timbral cues that identify a species, and this greatly facilitates communication. In speech communication, such processes may be responsible for what is referred to as vowel normalization (e.g. Miller 1989).

(a) *Communication sounds*

At the heart of each syllable of speech is a vowel. Figure 2 shows four versions of the vowel /a/ as in ‘hall’. From the auditory perspective, a vowel is a ‘pulse-resonance’ sound, that is, a stream of glottal pulses each with a resonance showing how the vocal tract responded to that pulse. From the speech perspective, the vowel contains three important components of the information in the larger communication (Irino & Patterson 2002). The first is the phonological ‘message’; for the vowels in figure 2, the message is that the vocal tract is currently in the shape that the brain associates with the phoneme /a/. This message is contained in the shape of the resonance which is the same in every cycle of all four waves. In figure 2*a*,*b* one person has spoken two versions of /a/ using a high and a low GPR, respectively; the pulse rate determines the pitch of the voice. The resonances have the same form since it is the same person speaking the same vowel. In figure 2*c*,*d* a short person and a tall person, respectively, have spoken versions of /a/ on the same pitch. The pulse rate and the shape of the resonance are the same, but the *rate* at which the resonance proceeds within the glottal cycle is slower in figure 2*d*. This person has the longer vocal tract and so their resonances ring longer. Since the vocal tract connects the mouth and nose to the lungs, VTL is highly correlated with the height of the speaker. In summary, it is the shape of the resonance that corresponds to the message or content of the speech sound. The GPR, which corresponds to the pitch, and the resonance rate, which corresponds to VTL, are derived from the ‘form’ of the message.

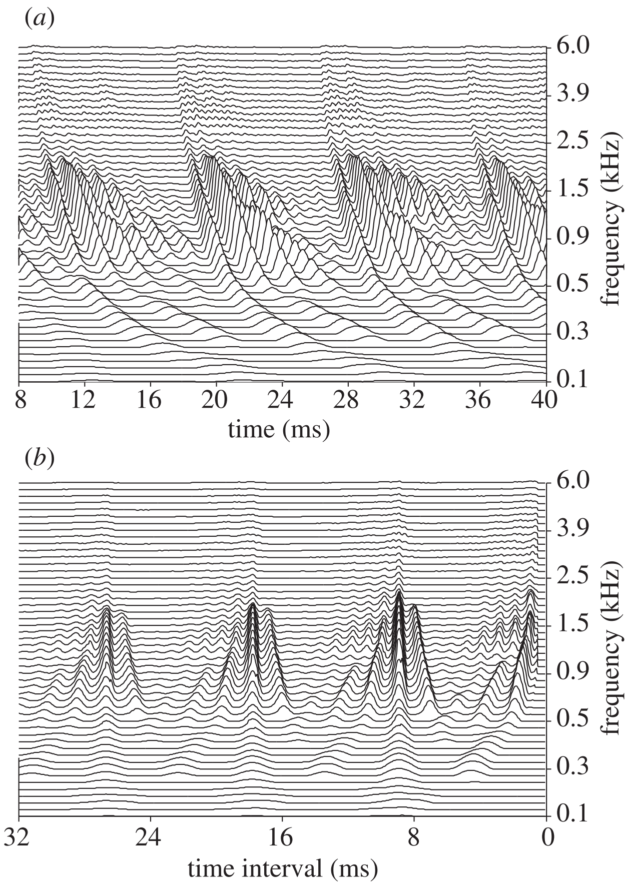

Figure 3. (*a*) The neural activity pattern and (*b*) the auditory image produced by the /a/ of ‘hat’. Note that the abscissa of the auditory image (*b*) is ‘time interval’ rather than time itself.

(b) *The auditory image model and auditory adaptation to GPR and VTL*

The general transforms involved in analysing GPR and VTL will be presented in the context of the auditory image model (AIM; Patterson *et al.* 1992, 1995), a model that focuses on the internal ‘auditory images’ produced by communication sounds and how these images can be produced with an ordered set of three transforms. The cochlea performs a spectral analysis of all incoming sounds. In AIM, this is simulated with an auditory filterbank in which the bandwidths of the filters are proportional to filter centre frequency. The filterbank converts an incoming sound wave into a multi-channel representation of basilar membrane motion. The most recent version of the auditory filter includes the fast-acting compression and two-tone suppression observed in the cochlea (Irino & Patterson 2006; Unoki *et al.* 2006). A ‘transduction’ mechanism involving half-wave rectification and low-pass filtering converts each channel of membrane motion into a simulation of the neural activity produced in the auditory nerve at that point on the basilar membrane. The result is a multi-channel, neural activity pattern (NAP) like that shown in figure 3*a*; the dimensions of the NAP are time (the abscissa) and auditory-filter centre frequency on a quasi-logarithmic axis (the ordinate). The surface defined by the set of lines is AIM’s simulation of the NAP produced in response to a short segment of this vowel. The channels cover the frequency range from 100 to 6000 Hz. The glottal pulses initiate activity in most of the channels every time they occur. The concentrations of energy in the mid-frequency region reveal the formants. Thus, the NAP of a vowel is a repeating pattern consisting of a warped vertical structure with triangular resonances on one side, which provide information about the shape of the vocal tract. The pattern repeats at the GPR which is heard as the voice pitch.

Whereas the activity in the NAP of a periodic sound oscillates on and off over the course of the glottal cycle, the percept evoked by such a stationary vowel does not flutter or wobble; indeed, periodic sounds produce the most stable of auditory perceptions. The contrast between the form of the NAP, which summarizes our understanding of the representation of sound in the early stages of the auditory pathway (CN and SOC), and the auditory image we hear indicate that there is some form of temporal integration between the NAP and the representation that is the basis of our initial perception of the sound. One process that could produce this stabilization is 'strobed' temporal integration (STI). It is assumed that there is a neural unit associated with each NAP channel that monitors its activity, to locate peaks like those produced by glottal pulses. The peaks cause the unit to 'strobe' the temporal integration process which (i) measures the time intervals from the strobe time to succeeding peaks in the decaying resonance and (ii) enters the time intervals into an interval histogram as they are generated (Patterson 1994). The histogram is dynamic; the information in it decays with a half-life of approximately 30 ms. The array of dynamic interval histograms across NAP channels is AIM's representation of the stabilized auditory image (SAI) that we hear in response to this kind of sound. The SAI of the NAP in figure 3*a* is presented in figure 3*b*. The rate of glottal cycles in speech is high relative to the rate of syllables; so, even for men who have GPRs in the range of 125 Hz, there are about four glottal cycles per 30 ms half-life. As a result, the level of the auditory image would be incremented four times during the time that it would otherwise take the image to decay to half its level, and so, a stable pattern builds up in the auditory image and remains there as long as the sound is on.

The process of stabilizing repeating patterns by calculating time intervals from NAP peaks applies to any periodic sound with amplitude modulation, that is, for sounds where one pulse of the period is somewhat larger than the others. This condition is common in the communication calls of animals and the sounds produced by most musical instruments (Fletcher & Rossing 1998; van Dinther & Patterson 2006). The normalization happens with the analysis; there is no need for a central pitch mechanism as in spectral models of perception (e.g. Terhardt 1974), and no need for a central neural net to learn that tokens of a vowel with different pitches should all be mapped to the same vowel type. The STI does provide information about pitch and pitch strength (Yost *et al.* 1996; Patterson *et al.* 2000), but the pitch information arises as a by-product of image stabilization and adaptation to the sound's pulse rate.

Adaptation to resonance rate, and thereby to VTL, involves a mathematically straightforward affine-scaling transform (Cohen 1993; Irino & Patterson 2002). When the VTL becomes shorter, the formants move up in frequency *and* they shrink in time. If the time-interval dimension in each channel of the auditory image is stretched, by multiplying time interval by the centre frequency of the channel, then the upper formants are stretched horizontally relative to the lower formants, and the mathematics tells us that the image becomes scale covariant. That is, changes in VTL just cause the formants to move up or down, as a group, without changing shape, and the vertical position of the pattern is the size information.

In summary, temporal models of auditory perception like AIM suggest that, following the initial frequency analysis performed in the cochlea, two relatively simple transforms are applied to the internal representation of the sound, which extract the pulse rate and resonance rate of the sound, and produce a largely invariant representation of the linguistic message. The result is that three kinds of information relating directly to GPR, VTL and to formant structure are available for processing. The model is helpful because it specifies a set of physiologically plausible processes and the order in which they occur. That is, the GPR and VTL adaptation mechanisms must occur, like frequency analysis and the binaural analysis of interaural phase and intensity cues, before the analysis of the sound's spectral structure (its timbre or formants). If the signature of one of the processes can be identified at a specific site in the auditory pathway, it places constraints on where the remaining processes are instantiated. This is the approach adopted by a loose consortium of auditory scientists to search initially for a 'hierarchy of pitch and melody processing' in the auditory system, and more recently to begin searching for the site of acoustic scaling in the auditory system. Neuroimaging research aimed at identifying and localizing the putative adaptation mechanisms is reviewed in §4.

4. IMAGING METHODS AND GENERAL AUDITORY PROCESSING

We began by searching for evidence of processing of periodicity, such as would be used to extract GPR, within the auditory pathway using functional magnetic resonance imaging (fMRI). fMRI is a non-invasive technique that indirectly measures regional neural activity, by measuring regional changes in blood oxygenation level. Griffiths *et al.* (2001) conducted a study with regular-interval (RI) sounds (Patterson *et al.* 1996; Yost *et al.* 1996) that are spectrally matched stimuli with and without temporal regularity, which give rise to a noisy percept with and without a buzzy pitch, respectively (see Patterson *et al.* 2002; figure 1). RI noise with pitch produces an auditory image with a vertical ridge at the pitch period, similar to the ridge produced by the glottal period in steady-state vowels, but without the formant structure. The study employed a 2-Tesla MR system, cardiac gating (Vlaardingerbroek & den Boer 1996; Guimaraes *et al.* 1998), sparse imaging (Edmister *et al.* 1999; Hall *et al.* 1999) and a magnet-compatible, high-fidelity sound system (Palmer *et al.* 1998). There were also 48 repetitions of each condition, which provided sufficient

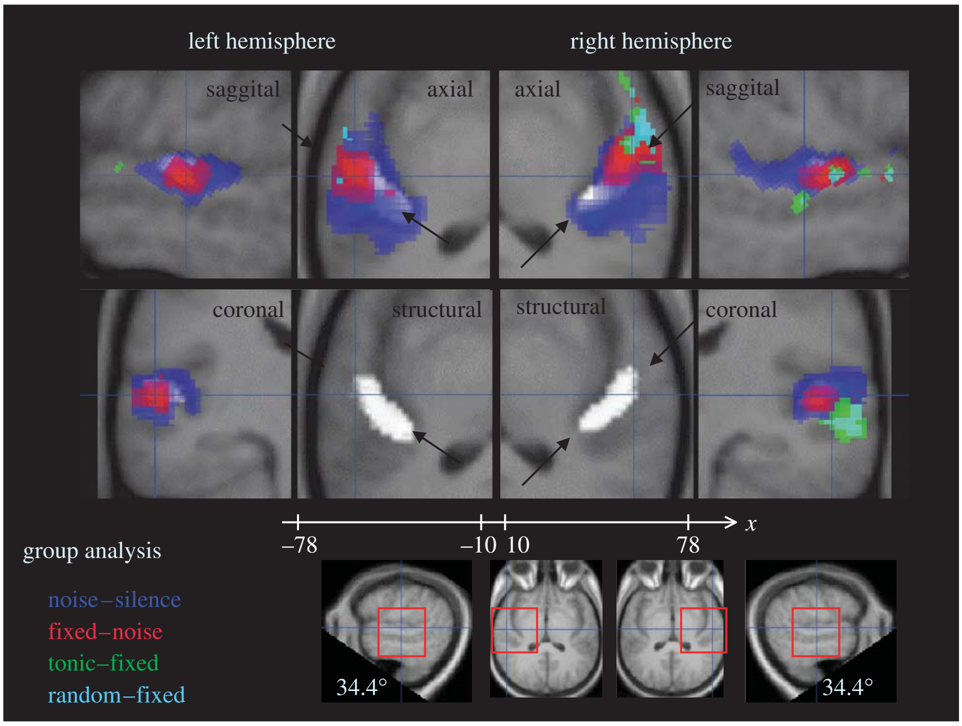

Figure 4. A summary of the results of Patterson *et al.* (2002). Group activation for four contrasts from Patterson *et al.* (2002), using a fixed-effects model, rendered onto the average structural image of the group (threshold $p<0.05$, corrected for multiple comparisons across the whole brain). The position and orientation of the sections are shown in the bottom panels of the figure. The axial sections show the activity in a plane parallel to the surface of the temporal lobe and just below it. The highlighted regions in the structural sections show the average position of HG in the two hemispheres; they are replotted under the functional activation in the axial sections above. The functional activation shows that, as a sequence of noise bursts acquires the properties of melody (first pitch and then changing pitch), the region sensitive to the added complexity changes from a large area on HG and planum temporale (blue), to a relatively focused area in the lateral half of HG (red), and then on out into surrounding regions of the planum polare (PP) and STG (green and cyan mixed). The orderly progression is consistent with the hypothesis that the hierarchy of melody processing that begins in the brainstem continues in auditory cortex and subsequent regions of the temporal lobe. The activation is largely symmetric in auditory cortex and becomes asymmetric abruptly as it moves on to PP and STG with relatively more activity in the right hemisphere.

sensitivity to reveal activation in the four major subcortical nuclei of the auditory pathway. The contrast between the activation produced by sounds with and without temporal regularity revealed that the processing of temporal regularity begins in subcortical structures (CN and IC). A contrast with the same power, between sounds with a fixed pitch and sounds where the pitch was varied to produce a melody, revealed that changing pitch does not produce more activation than fixed pitch in these nuclei.

The cortical activation from this fMRI study was reported in Patterson *et al.* (2002); the main results are presented in figure 4 (fig. 3, Patterson *et al.* 2002). The morphological landmark for PAC in humans is HG (Rademacher *et al.* 1993; Rivier & Clarke 1997; Morosan *et al.* 2001; Rademacher *et al.* 2001). The location of HG was identified in each of the subjects; the conjoint volume is shown in white in the central panels of figure 4 and its location is in good agreement with the locations reported in other studies (Penhune *et al.* 1996; Leonard *et al.* 1998). The figure shows that noise on its own (the blue regions in figure 4) produced more activation than silence, bilaterally, in a large cluster of voxels centred on HG and planum temporale (PT) behind it. The same region is activated by the stimuli with temporal regularity, whether the regularity is fixed (so that pitch is fixed) or varying (as in the melodies). The activation peaks in this region are highly significant and they appear with remarkable consistency in all of the contrasts between sound and silence, and in all subjects.

The region of the noise—silence contrast (blue) at the medial end of HG was particularly consistent. In this region, despite strong activation to all stimuli, there was no differential activity when activation to one sound condition was contrasted with that of any other. For example, figure 4 shows that when the fixed-pitch condition was contrasted with noise (red), or when the changing-pitch conditions (melody) were contrasted with noise (cyan and green), there was no differential activation in medial HG. The obvious interpretation is that this is core auditory cortex (PAC) which is fully engaged by the processing of any complex sound, so the level of activation is the same for sounds with the same loudness.

When the fixed-pitch condition was contrasted with noise (red), it revealed differential activation, bilaterally, in anterolateral HG (al-HG), a region that could be auditory belt (Wallace *et al.* 2002) or core (Morosan *et al.* 2001). When activity in the melody conditions was contrasted with that in the fixed-pitch condition, it revealed differential activation in planum polare (PP) and in STG just below HG (cyan and green in figure 4), and the activation was stronger in the right hemisphere. In bilateral al-HG, melody conditions produced roughly the same level of activity as fixed-pitch conditions. Taken together, these results suggest that neurons in the al-HG region are involved in the cross-frequency evaluation of pitch value and pitch strength, and subsequent temporal regions, particularly in the right hemisphere, evaluate changes in pitch over time.

Penagos *et al.* (2004) extended the results of Griffiths *et al.* (2001) and Patterson *et al.* (2002) using harmonic complex tones with and without resolved harmonics. They used a 3-Tesla MR system, cardiac gating and sparse imaging which enabled them to measure activation in CN, IC and HG, and show that the level of activation does not vary significantly with pitch salience, except in a small region of al-HG, bilaterally. There were no changing-pitch conditions in this study. Finally, Bendor & Wang (2005) have recently reported finding cells in marmoset (*Callithrix jacchus*) cortex sensitive to the low pitch of harmonic complex tones. The cells were in an auditory core area adjacent to A1 (area R) which, Bendor & Wang (2005) argue, is probably homologous to al-HG in humans.

The results of these experiments can be interpreted to indicate that there is a 'hierarchy of processing' in the auditory pathway. With regard to AIM and the transform that adapts to GPR, the results suggest that (i) the extraction of time-interval information from the firing pattern in the auditory nerve probably occurs in the brainstem, (ii) the construction of the time-interval histograms probably occurs in, or near, the thalamus (MGB), and the resulting SAI is in PAC, (iii) the cross-channel evaluation of pitch value and pitch strength probably occurs in al-HG, and (iv) assessment of pitch variation for the perception of melody, and perhaps prosody appears to occur in regions beyond auditory cortex (anterior STG) particularly in the right hemisphere. It is this last process, requiring integration over long time periods that gives rise to the hemispheric asymmetries observed in neuropsychological and functional neuroimaging studies of pitch perception, rather than pitch extraction *per se* (Patterson *et al.* 2002)

(a) *Imaging the auditory system with MEG*

MEG measures the strength and direction of the magnetic dipole produced by activation in nerve fibres running parallel to the scalp; it is largely insensitive to radial sources. This is an advantage for measuring activity along the surface of the temporal lobe in the region of lateral belt and parabelt auditory cortex, and a disadvantage for imaging of activity in higher-level auditory areas like the STS. However, the main advantage of MEG for the investigation of auditory function is that it has millisecond temporal resolution which can be used to investigate the order of events in auditory cortex. It is also the case that recent MEG machines with hundreds of sensors make it possible to localize sources sufficiently well to associate them with regions of activation observed with fMRI.

The auditory evoked field (AEF) is dominated by a large negative deflection associated with stimulus onset; it appears in the interval between 80 and 130 ms post-stimulus onset, and when the source can be located, it is usually in PT just posterior to al-HG. It is referred to as N1m or N100m, and it is generally assumed to represent the aggregate activity of several sources involved in general auditory processing. There are several techniques for dissecting the components of this large, broad, negative deflection. One can simply gather sufficient MEG data to reveal smaller positive and negative peaks on the flanks of the N100m by averaging. For example, Lütkenhöner *et al.* (2003) showed that the first cortical response in humans to transient sounds is a negative deflection in PAC 19 ms post-stimulus onset (N19m), and Rupp *et al.* (2002) showed that short chirps produce an N19m–P30m complex from a source on HG near PAC. The P30 is the first of a set of positive field generators that appear in, or near, PAC between 30 and 60 ms post-onset, and which are collectively referred to as 'P1m'. Gutschalk *et al.* (2004) have shown that the P1m is related to stimulus onset but not the processing of temporal regularity.

Forss *et al.* (1993) have shown that the latency of the N100m elicited by a regular click train is inversely related to the pitch of the sound, which led Crottaz-Herbette & Ragot (2000) to propose that the generators of the N100m are involved in pitch processing. However, an earlier review of a wide range of studies (Näätänen & Picton 1987) concluded that an N100m can be elicited by the onset of almost any kind of sound. So, while it is the case that the latency of the N100m varies with pitch, the response is fundamentally confounded with the activation of other generators that reflect features like loudness and timbre rather than pitch. To isolate the pitch component of the N100m, Krumbholz *et al.* (2003) developed a continuous stimulation technique in which the sound begins with a stationary noise and then, after a second or so, when the N100m has passed and the AEF has settled into a sustained response, the fine structure of the noise is regularized to produce a RI sound (RIS) without changing the energy or spectral distribution of the energy. There is a marked perceptual change at the transition from noise to RIS, and it is accompanied by a prominent negative deflection in the magnetic field, referred to as the pitch onset response (POR). The inverse transition, from RIS to noise, produces virtually no deflection. Krumbholz *et al.* (2003) showed that the latency of the POR varies inversely with the pitch of the RIS, and the magnitude of the response increases with pitch strength. Gutschalk *et al.* (2004) constructed a continuous stimulus of alternating regular and irregular click trains, and used it to isolate the POR from an intensity-related response in PT. The source of the POR was located in al-HG very near the pitch centre identified by Patterson *et al.* (2002) and Penagos *et al.* (2004). The PORs were surprisingly late: approximately 120 ms *plus* four times the period of the click train. This is substantially longer than might be anticipated from temporal models of pitch perception (e.g. Patterson *et al.* 1995; Krumbholz *et al.* 2003). The results suggest that the POR reflects relatively late cortical processes involved in the cross-channel estimation of pitch and pitch strength. It stands in sharp contrast to the initial extraction of periodicity information with STI which is thought to be in the brainstem and thalamus.

The notes of music and the vowels of speech produce sustained pitch perceptions, and there is a sustained component of the AEF that appears to reflect the sustained perception of pitch. The advent of MEG systems with 125–250 sensors means that it is now possible to measure sustained fields. Gutschalk *et al.* (2002) contrasted the activity produced by regular and irregular click trains and performed the experiment at three intensities. This enabled them to isolate two sources in each hemisphere adjacent to PAC. The more anterior source was located in lateral HG in the pitch region identified with fMRI by Patterson *et al.* (2002) and Penagos *et al.* (2004). This source was particularly sensitive to regularity and largely insensitive to sound level. The second source was located just posterior to the first in PT; it was particularly sensitive to sound level and largely insensitive to regularity. This double dissociation provided convincing evidence that the source of the POR in al-HG also produces a sustained field that is related to the sustained perception of pitch. The posterior source in PT would appear to be more involved with the perception of loudness.

The studies discussed to this point are all related to the transforms that adapt auditory analysis to the GPR of the vowel and evaluate the pitch. There is only one very recent study (von Kriegstein *et al.* 2006) of the processes that adapt auditory analysis to the resonance rate of the vowel, that is, the VTL of the speaker. Subjects listened to sequences of syllables in which GPR and VTL either remained fixed or varied randomly in a 2×2 factorial design. The results are compatible with the model of hierarchical processing inasmuch they indicate that the adaptation begins in the MGB and is not completed until regions of STG beyond auditory cortex. However, these initial results do not reveal one simple region for the processing of VTL information. The cortical activation arises from the interaction of GPR and VTL, which occurs naturally as children grow; however, it complicates the interpretation of the results.

5. THE BEGINNINGS OF VOICE-SPECIFIC AND SPEECH-SPECIFIC PROCESSING IN THE BRAIN, AND THE IMPLICATIONS FOR MODELS OF SPEECH PERCEPTION

As noted above, talker normalization is an important preliminary step to recovering the content of an utterance. Not until GPR and VTL have been taken into account, and a transformed auditory image reflecting the shape of the vocal apparatus achieved, can the linguistic content (phonemes/words/phrases) be analysed and interpreted. Until recently, research on speech perception (as a linguistic signal) assumed that talker-specific information, such as GPR and VTL, was simply stripped away from the linguistic content of the message relatively early in processing—it was thought that this was simply noise that did not contribute helpfully to interpreting speech. A growing body of research makes it clear, however, that form and content must, to some extent, be processed together (Goldinger 1998). Detailed information about an individual talker's voice is encoded and retained and can subsequently improve intelligibility of this familiar voice (Pisoni 1997; Nygaard & Pisoni 1998; Sheffert *et al.* 2002; Bent & Bradlow 2003). Thus, the transformations discussed in §3 must also permit talker-specific information such as GPR and VTL to contact the linguistic information processed in cortex.

The brain networks underlying both voice-specific processing and the transformation from general auditory to speech-specific processing have been studied using imaging methods, particularly fMRI. Voice-specific processing has not been extensively investigated, and speech-specific processing—particularly the question of where the transformation from auditory to linguistic processing occurs—has been neglected until recently. In fact, this is not one but many questions: Is the transformation localizable? If localizable, do we observe evidence for one neural locus or for several? Is such a transformation dependent only upon the acoustics of the signal or also upon the cognitive state of the individual? Is this a modular auditory process, or can it be influenced by other sensory modalities (i.e. visual; somatomotor)? Recent imaging studies have begun to answer these questions. First, however, we will briefly review the evidence supporting localized voice-specific processing in cortex.

Thierry *et al.* (2003) compared auditory processing of speech and environmental sound sequences matched for duration, rhythm, content and interpretability, in addition to identifying a network of areas activated during comprehension of both kinds of sound, several areas in the left anterior superior and middle temporal gyri, straddling the STS, were activated more by speech than by environmental sounds. As the authors point out (see Price *et al.* 2005), increased sensitivity to speech over environmental sounds could arise either because these STS areas are sensitive to speech *qua* linguistic signal or *qua* vocal signal. It is difficult to separate these two types of processing, but studies indicate that both types of processing appear to recruit similar regions, in both hemispheres. However, the processing of speech as a linguistic signal seems to recruit left-hemisphere STS areas preferentially, whereas processing of speech as a voice signal seems to recruit right-hemisphere STS areas preferentially.

Pascal Belin and colleagues (Belin *et al.* 2000, 2002; Fecteau *et al.* 2004; see Belin *et al.* 2004 for a review) used non-speech vocalizations such as laughs and cries to distinguish between processing of speech as voice and speech as linguistic content. They observed robust bilateral activity in the STS for human voices when compared with non-human sounds, irrespective of the linguistic content of the voice (Belin *et al.* 2000, 2002). In most listeners the peak of this activity appeared to be in the upper bank of the STS. Although anatomical homologies between humans and macaque monkeys in the STS are not yet known, this region in the macaque corresponds to a third or fourth stage of cortical auditory processing (Kaas *et al.* 1999, fig. 2*a*), and may similarly subserve late-stage auditory processing in humans. In a subsequent study with different listeners, similar, but somewhat more inferior, regions exhibited greater sensitivity to human than to animal vocalizations, suggesting that these STS regions are not just sensitive to voices, but are more sensitive to human voices. Von Kriegstein & Giraud (2004) used two different tasks with the same set of sentences spoken by different talkers to observe three regions in the right STS that are more active during recognition of target voices than during recognition of target sentence content, confirming a role for multiple right STS regions in voice perception. Unlike the studies by Belin and colleagues, differential activity in left STS regions was not observed in this contrast. Note however that voices were equally present in all conditions: if voice processing is more obligatory in the left hemisphere (perhaps because it a necessary concomitant of the extraction of linguistic content), then it would be 'subtracted out' in the contrast.

In general, the evidence suggests that the transformation from an auditory signal to speech is localizable and is distributed across several neural loci, including PT (probable belt or parabelt) and STS but not HG (probable core and belt). Uppenkamp *et al.* (2006) scanned volunteers while they listened to natural and synthetic vowels, or to non-speech stimuli matched to the vowel sounds in terms of their long-term energy and spectro-temporal profiles. Vowels produced more activation than non-speech sounds in several regions along the STS bilaterally, in anterolateral PT as well as in premotor cortex, but not in any anatomically earlier auditory region Other researchers observe multiple foci throughout the temporal lobe when speech and non-speech perceptions are compared, even despite the 8–12 mm smoothing that is common in imaging studies (e.g. Giraud *et al.* 2004; Rimol *et al.* 2005). Jacquemot *et al.* (2003) performed a study in which they examined the neural correlates of acoustic differences within, or across, phonetic categories. They exploited cross-linguistic differences in phonology between French and Japanese, to achieve a counterbalanced design in which stimuli that were perceived by one language group as belonging to the same phonetic category were perceived by the other group as belonging to different phonetic categories. When across- versus within-category stimuli were compared across groups, activation was observed in the supramarginal gyrus and in a region of anterior PT. Since linguistic stimuli were present in both conditions, any activation that was due to these stimuli being treated as speech would be subtracted out, and the activity could reflect some common, experience-dependent, magnification of acoustic differences across groups. The results are generally compatible with the hierarchical model of primate auditory processing, and with the idea that early cortical stages of processing respond indiscriminately to speech and non-speech sounds, and only regions at a higher stage of processing are specialized for speech perception. However, even though speech and non-speech stimuli were acoustically closely matched in the study by Uppenkamp *et al.* (2006), the synthetic speech stimuli had a perceptual coherence, eliciting a voice-like percept that the non-speech stimuli lacked. Thus, we cannot determine whether the activation foci observed in this study, and in many other studies comparing speech and non-speech stimuli (e.g. Demonet *et al.* 1992; Binder *et al.* 1997, 2000; Jancke *et al.* 2002), reflect voice or linguistic perception, or both.

Activity in speech-sensitive regions can be contingent upon the cognitive state of the individual. In several fMRI experiments (Liebenthal *et al.* 2003; Giraud *et al.* 2004; Dehaene-Lambertz *et al.* 2005; Möttönen *et al.* 2006) the physical identity of auditory stimuli was held constant between two conditions, but the cognitive state of the listeners was systematically manipulated so that they heard the stimuli as non-speech in one condition, but as speech in another. For example, in the study by Giraud *et al.* (2004), listeners heard sentences that had been noise-vocoded (divided into four frequency bands and then resynthesized onto a noise carrier) both before and after a period of training. These sentences were initially not heard as speech, but after they had been presented pairwise with their natural-speech homologues (training), listeners could understand them. A control stimulus set consisted of vocoded sentences that were degraded acoustically so that training did not increase comprehension. The post- versus pre-training contrast was essentially tested as an interaction with post–pre control stimuli, to remove systematic order-of-testing effects, and revealed areas that are selectively active during comprehension. Interestingly, this contrast did not reveal activity in the left STS, instead activity was observed in right STS and in left and right middle and inferior temporal areas. Dehaene-Lambertz *et al.* (2005) used sine-wave consonant–vowel syllables which can be perceived either as non-speech whistles or, following instructions and training, as syllables. Posterior left STS/STG was the only region that was more active when listeners heard the stimuli as speech compared to when they heard them as non-speech. Möttönen *et al.* (2006) also demonstrated left posterior STS activity when perceiving sine-wave processed non-word bisyllables as speech compared to non-speech. Again, whether such activation arises as a result of voice or linguistic perception is an open question.

Researchers have observed activity in premotor cortex or motor cortex during the perception of speech (words: Fadiga *et al.* 2002; monosyllables: Wilson *et al.* 2004, Uppenkamp *et al.* 2006; connected speech: Watkins *et al.* 2003, Watkins & Paus 2004). It is therefore possible that the acoustic-to-speech transformation relies on multiple regions in a distributed network including both temporal-lobe and motor–premotor regions, although the stimuli used in these studies may have engaged lexical, semantic and syntactic processes in addition to speech-sound processing, and further studies are required to determine whether premotor/motor activity reflects processes that are prerequisite to speech perception, or instead reflects some late process that is merely correlated with speech perception (e.g. semantically relevant imagery; cf. Hauk *et al.* 2004).

The anatomical connections between the auditory system and motor structures are highly compatible with a wealth of information attesting to speech perception as a sensorimotor phenomenon. Several models of speech perception, positing a basis in articulatory or gestural representations, have been formulated (Fowler 1996; Rizzolatti & Arbib 1998; Liberman & Whalen 2000; MacNeilage & Davis 2001). The motor theory of speech perception is unlikely to hold in its orthodox form (e.g. Liberman & Whalen 2000), but more moderate positions that acknowledge at least preliminary domain-general auditory processing of the speech signal also propose that speech perception and production are linked (Kluender & Lotto 1999). This is seen most clearly during development where imitation plays an important role in children's acquisition of spoken language (Kuhl 1994; Doupe & Kuhl 1999) but sensorimotor development would have an impact on the organization of speech perception in the adult brain.

Most authors have concluded that such motor activity during speech perception could reflect the activation of articulatory gestural representations which permit the listener to derive the *intended gesture* of the speaker; this is in line with the motor theory of speech perception (e.g. Liberman & Whalen 2000). Such access to gestural representations provides for parity; a shared code between speaker and listener, which is essential for speech communication (Rizzolatti & Arbib 1998; Liberman & Whalen 2000).

The highly parallel organization of cortical regions and their interconnections suggests a way that auditory and gestural accounts of speech perception can be reconciled (Scott & Johnsrude 2003). Different processing pathways may be differentially specialized to serve different processes or operate on complementary representations of speech (phonological versus articulatory). Multiple pathways, operating in parallel, would serve to make speech perception the efficient and robust communication system we know it to be.

6. CONCLUSIONS

The incoming auditory signal is extensively processed and recoded by the time it reaches auditory cortex, and it probably is not treated in any speech-specific way until relatively late—atleast three or four cortical processing stages beyond PAC. Prior to that stage, processes are more domain general—more about the form of the speech (how the talker was talking) and less about the content of the speech (what the talker was saying). These two classes of processing work together to recover the speech content and information indicative of the size, sex and age of the talker among other indexical characteristics. The mammalian auditory system is organized hierarchically, and from PAC onwards the anatomy suggests multiple, parallel, processing systems with strong feedback connections suggesting that multiple aspects of the speech signal are processed more-or-less simultaneously with reference to the ongoing context. The fact that the processing is distributed means that functional imaging (fMRI and MEG) can assist in exploring both the subcortical and cortical networks involved in domain-general and speech-specific processing, and the interactions among them.

The writing of this paper was supported by a grant to the first author from the UK Medical Research Council (G0500221).

REFERENCES

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. & Zilles, K. 1999 Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* **412**, 319–341. (doi:10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7)

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. 2000 Voice-selective areas in human auditory cortex. *Nature* **403**, 309–312. (doi:10.1038/35002078)

Belin, P., Zatorre, R. J. & Ahad, P. 2002 Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* **13**, 17–26. (doi:10.1016/S0926-6410(01)00084-2)

Belin, P., Fecteau, S. & Bedard, C. 2004 Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* **8**, 129–135. (doi:10.1016/j.tics.2004.01.008)

Bendor, D. & Wang, X. 2005 The neuronal representation of pitch in primate auditory cortex. *Nature* **436**, 1161–1165. (doi:10.1038/nature03867)

Bent, T. & Bradlow, A. R. 2003 The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.* **114**, 1600–1610. (doi:10.1121/1.1603234)

Binder, J. R., Frost, J., Hammeke, T., Cox, R., Rao, S. & Prieto, T. 1997 Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* **17**, 353–362.

Binder, J. R., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J. & Possing, E. 2000 Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528. (doi:10.1093/cercor/10.5.512)

Brugge, J. F., Volkov, I., Garell, P., Reale, R. & Howard III, M. 2003 Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J. Neurophysiol.* **90**, 3750–3763. (doi:10.1152/jn.00500.2003)

Cohen, L. 1993 The scale transform. *IEEE Trans. Acoust. Speech Signal Process.* **41**, 3275–3292.

Crottaz-Herbette, S. & Ragot, R. 2000 Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clin. Neurophysiol.* **111**, 1759–1766. (doi:10.1016/S1388-2457(00)00422-3)

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A. & Dehaene, S. 2005 Neural correlates of switching from auditory to speech perception. *NeuroImage* **24**, 21–33. (doi:10.1016/j.neuroimage.2004.09.039)

Demonet, J., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J., Wise, R., Rascol, A. & Frackowiak, R. 1992 The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768. (doi:10.1093/brain/115.6.1753)

Doupe, A. & Kuhl, P. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)

Edmister, W., Talavage, T., Ledden, P. & Weisskoff, R. 1999 Improved auditory cortex imaging using clustered volume acquisitions. *Hum. Brain. Mapp.* **7**, 89–97. (doi:10.1002/(SICI)1097-0193(1999)7:2<89::AID-HBM2>3.0.CO;2-N)

Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. 2002 Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* **15**, 399–402. (doi:10.1046/j.0953-816x.2001.01874.x)

Fecteau, S., Armony, J., Joanette, Y. & Belin, P. 2004 Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage* **23**, 840–848. (doi:10.1016/j.neuroimage.2004.09.019)

Fletcher, N. H. & Rossing, T. 1998 *The physics of musical instruments*. New York, NY: Springer.

Forss, N., Makela, J., McEvoy, L. & Hari, R. 1993 Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains. *Hear. Res.* **68**, 89–96. (doi:10.1016/0378-5955(93)90067-B)

Fowler, C. A. 1996 Listeners do hear sounds, not tongues. *J. Acoust. Soc. Am.* **99**, 1730–1741. (doi:10.1121/1.415237)

Garell, P., Volkov, I., Noh, M., Damasio, H., Reale, R., Hind, J., Brugge, J. & Howard, M. 1998 Electrophysiologic connections between the posterior superior temporal gyrus and lateral frontal lobe in humans. *Soc. Neurosci. Abstr.* **24**, 1877.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M., Preibisch, C. & Kleinschmidt, A. 2004 Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb. Cortex* **14**, 247–255. (doi:10.1093/cercor/bhg124)

Goldinger, S. D. 1998 Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* **105**, 251–279. (doi:10.1037/0033-295X.105.2.251)

Greenlee, J. D., Oya, H., Kawasaki, H., Volkov, I., Kaufman, O., Kovach, C., Howard, M. & Brugge, J. 2004 A functional connection between inferior frontal gyrus and orofacial motor cortex in human. *J. Neurophysiol.* **92**, 1153–1164. (doi:10.1152/jn.00609.2003)

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O. & Patterson, R. D. 2001 Encoding of the temporal regularity of sound in the human brainstem. *Nat. Neurosci.* **4**, 633–637. (doi:10.1038/88459)

Guimaraes, A. R., Melcher, J., Talavage, T., Baker, J., Ledden, P., Rosen, B., Kiang, N., Fullerton, B. & Weisskoff, R. 1998 Imaging subcortical auditory activity in humans. *Hum. Brain Mapp.* **6**, 33–41. (doi:10.1002/(SICI)1097-0193(1998)6:1<33::AID-HBM3>3.0.CO;2-M)

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S. & Scherg, M. 2002 Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage* **15**, 207–216. (doi:10.1006/nimg.2001.0949)

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S. & Rupp, A. 2004 Temporal dynamics of pitch in human auditory cortex. *NeuroImage* **22**, 755–766. (doi:10.1016/j.neuroimage.2004.01.025)

Hackett, T. A., & Kaas, J. H. 2004 Auditory cortex in primates: functional subdivisions and processing streams. In *The cognitive neurosciences III* (ed. M. Gazzaniga), ch. 16, pp. 215–232. Cambridge, MA: MIT Press.

Hackett, T. A., Stepniewska, I. & Kaas, J. 1998 Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* **394**, 475–495. (doi:10.1002/(SICI)1096-9861(19980518)394:4<475::AID-CNE6>3.0.CO;2-Z)

Hackett, T. A., Stepniewska, I. & Kaas, J. 1999 Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* **817**, 45–58. (doi:10.1016/S0006-8993(98)01182-2)

Hackett, T. A., Preuss, T. & Kaas, J. 2001 Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* **441**, 197–222. (doi:10.1002/cne.1407)

Hall, D. A., Haggard, M., Akeroyd, M., Palmer, A., Summerfield, A. Q., Elliott, M., Gurney, E. & Bowtell, R. 1999 "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* **7**, 213–223. (doi:10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N)

Hall, D. A., Hart, H. & Johnsrude, I. 2003 Relationships between human auditory cortical structure and function. *Audiol. Neurootol.* **8**, 1–18. (doi:10.1159/000067894)

Hauk, O., Johnsrude, I. & Pulvermuller, F. 2004 Somatotopic representation of action words in human motor and premotor cortex. *Neuron* **41**, 301–307. (doi:10.1016/S0896-6273(03)00838-9)

Howard, M. *et al.* 2000 Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* **416**, 79–92. (doi:10.1002/(SICI)1096-9861(20000103)416:1<79::AID-CNE6>3.0.CO;2-2)

Irino, T. & Patterson, R. D. 2002 Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-Mellin transform. *Speech Commun.* **36**, 181–203. (doi:10.1016/S0167-6393(00)00085-6)

Irino, T. & Patterson, R. D. 2006 A dynamic, compressive gammachirp auditory filterbank. *IEEE Audio, Speech Lang. Process (ASLP)* **14**, 2222–2232. (doi:10.1109/TASL.2006.874669)

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S. & Dupoux, E. 2003 Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* **23**, 9541–9546.

Jancke, L., Wustenberg, T., Scheich, H. & Heinze, H. 2002 Phonetic perception and the temporal cortex. *NeuroImage* **15**, 733–746. (doi:10.1006/nimg.2001.1027)

Jones, E. G., Dell'Anna, M., Molinari, M., Rausell, E. & Hashikawa, T. 1995 Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J. Comp. Neurol.* **362**, 153–170. (doi:10.1002/cne.903620202)

Kaas, J. & Hackett, T. 2000 Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl Acad. Sci. USA* **97**, 11 793–11 799. (doi:10.1073/pnas.97.22.11793)

Kaas, J., Hackett, T. & Tramo, M. 1999 Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* **9**, 164–170. (doi:10.1016/S0959-4388(99)80022-1)

Kluender, K. R. & Lotto, A. 1999 Virtues and perils of an empiricist approach to speech perception. *J. Acoust. Soc. Am.* **105**, 503–511. (doi:10.1121/1.424587)

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C. & Lütkenhöner, B. 2003 Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cereb. Cortex* **13**, 765–772. (doi:10.1093/cercor/13.7.765)

Kuhl, P. 1994 Learning and representation in speech and language. *Curr. Opin. Neurobiol.* **4**, 812–822. (doi:10.1016/0959-4388(94)90128-7)

Leonard, C., Puranik, C., Kuldau, J. & Lombardino, L. 1998 Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? *Cereb. Cortex* **8**, 397–406. (doi:10.1093/cercor/8.5.397)

Liberman, A. & Whalen, D. 2000 On the relation of speech to language. *Trends Cogn. Sci.* **4**, 187–196. (doi:10.1016/S1364-6613(00)01471-6)

Liebenthal, E., Binder, J., Piorkowski, R. & Remez, R. 2003 Short-term reorganization of auditory analysis induced by phonetic experience. *J. Cogn. Neurosci.* **15**, 549–558. (doi:10.1162/089892903321662930)

Liegeois-Chauvel, C., Musolino, A. & Chauvel, P. 1991 Localization of the primary auditory area in man. *Brain* **114**, 139–151.

Lütkenhöner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinstrater, O. & Patterson, R. D. 2003 Localization of primary auditory cortex in humans by magnetoencephalography. *NeuroImage* **18**, 58–66. (doi:10.1006/nimg.2002.1325)

MacNeilage, P. F. & Davis, B. L. 2001 Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Curr. Opin. Neurobiol.* **11**, 696–700. (doi:10.1016/S0959-4388(01)00271-9)

Merzenich, M. & Brugge, J. 1973 Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* **50**, 275–296. (doi:10.1016/0006-8993(73)90731-2)

Miller, J. D. 1989 Auditory–perceptual interpretation of the vowel. *J. Acoust. Soc. Am.* **85**, 2114–2134. (doi:10.1121/1.397862)

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T. & Zilles, K. 2001 Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* **13**, 684–701. (doi:10.1006/nimg.2000.0715)

Möttönen, R., Calvert, G., Jääskeläinen, I., Matthews, P., Thesen, T., Tuomainen, J. & Sams, M. 2006 Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage* **30**, 563–569. (doi:10.1016/j.neuroimage.2005.10.002)

Näätänen, R. & Picton, T. 1987 The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* **24**, 375–425. (doi:10.1111/j.1469-8986.1987.tb00311.x)

Nygaard, L. & Pisoni, D. 1998 Talker-specific learning in speech perception. *Percept. Psychophys.* **60**, 355–376.

Palmer, A. R., Bullock, D. & Chambers, J. 1998 A high-output, high-quality sound system for use in auditory fMRI. *NeuroImage* **7**, S359.

Pandya, D. N. 1995 Anatomy of the auditory cortex. *Rev. Neurol. (Paris)* **151**, 486–494.

Pandya, D. N. & Sanides, F. 1973 Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Z. Anat. Entwicklungsgesch.* **139**, 127–161. (doi:10.1007/BF00523634)

Pandya, D. N. & Seltzer, B. 1982 Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey. *J. Comp. Neurol.* **204**, 196–210. (doi:10.1002/cne.902040208)

Patterson, R. D. 1994 The sound of a sinusoid: time-interval models. *J. Acoust. Soc. Am.* **96**, 1419–1428. (doi:10.1121/1.410286)

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. & Allerhand, M. 1992 Complex sounds and auditory images. In *Auditory physiology and perception, Proc. 9th Int. Symp. Hear.* (eds Y. Cazals, L. Demany & K. Horner), pp. 429–446. Oxford, UK: Pergamon.

Patterson, R. D., Allerhand, M. & Giguere, C. 1995 Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* **98**, 1890–1894. (doi:10.1121/1.414456)

Patterson, R. D., Handel, S., Yost, W. & Datta, A. 1996 The relative strength of the tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.* **100**, 3286–3294. (doi:10.1121/1.417212)

Patterson, R. D., Hackney, C. M. & Iversen, I. D. 1999 Interdisciplinary auditory neuroscience. *Trends Cognit. Neurosci.* **3**, 245–247. (doi:10.1016/S1364-6613(99)01347-9)

Patterson, R. D., Yost, W., Handel, S. & Datta, A. 2000 The perceptual tone/noise ratio of merged iterated rippled noises. *J. Acoust. Soc. Am.* **107**, 1578–1588. (doi:10.1121/1.428442)

Patterson, R. D., Uppenkamp, S., Johnsrude, I. & Griffiths, T. 2002 The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767–776. (doi:10.1016/S0896-6273(02)01060-7)

Penagos, H., Melcher, J. & Oxenham, A. 2004 A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* **24**, 6810–6815. (doi:10.1523/JNEUROSCI.0383-04.2004)

Penhune, V. B., Zatorre, R. J., MacDonald, J. & Evans, A. 1996 Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* **6**, 661–672. (doi:10.1093/cercor/6.5.661)

Petrides, M. & Pandya, D. N. 1984 Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* **228**, 105–116. (doi:10.1002/cne.902280110)

Petrides, M. & Pandya, D. N. 1988 Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* **273**, 52–66. (doi:10.1002/cne.902730106)

Petrides, M. & Pandya, D. N. 2002 Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* **16**, 291–310. (doi:10.1046/j.1460-9568.2001.02090.x)

Pisoni, D. 1997 Some thoughts on 'normalization' in speech perception. In *Talker variability in speech processing* (eds K. Johnson & J. W. Mullennix), pp. 9–32. San Diego, CA: Academic Press.

Price, C., Thierry, G. & Griffiths, T. 2005 Speech-specific auditory processing: where is it? *Trends Cogn. Sci.* **9**, 271–276. (doi:10.1016/j.tics.2005.03.009)

Rademacher, J., Caviness Jr, V., Steinmetz, H. & Galaburda, A. 1993 Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cereb. Cortex* **3**, 313–329. (doi:10.1093/cercor/3.4.313)

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. & Zilles, K. 2001 Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* **13**, 669–683. (doi:10.1006/nimg.2000.0714)

Rauschecker, J. P. 1998 Parallel processing in the auditory cortex of primates. *Audiol. Neurootol.* **3**, 86–103. (doi:10.1159/000013784)

Rauschecker, J. P. & Tian, B. 2000 Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806. (doi:10.1073/pnas.97.22.11800)

Rauschecker, J. P., Tian, B., Pons, T. & Mishkin, M. 1997 Serial and parallel processing in rhesus monkey auditory cortex. *J. Comp. Neurol.* **382**, 89–103. (doi:10.1002/(SICI)1096-9861(19970526)382:1<89::AID-CNE6>3.0.CO;2-G)

Rimol, L. M., Specht, K., Weis, S., Savoy, R. & Hugdahl, K. 2005 Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *NeuroImage* **26**, 1059–1067. (doi:10.1016/j.neuroimage.2005.03.028)

Rivier, F. & Clarke, S. 1997 Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex. *NeuroImage* **6**, 288–304. (doi:10.1006/nimg.1997.0304)

Rizzolatti, G. & Arbib, M. 1998 Language within our grasp. *Trends Neurosci.* **21**, 188–194. (doi:10.1016/S0166-2236(98)01260-0)

Romanski, L., Bates, J. & Goldman-Rakic, P. 1999*a* Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* **403**, 141–157. (doi:10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V)

Romanski, L., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. & Rauschecker, J. 1999*b* Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* **2**, 1131–1136. (doi:10.1038/16056)

Rupp, A., Uppenkamp, S., Gutschalk, A., Beucker, R., Patterson, R. D., Dau, T. & Scherg, M. 2002 The representation of peripheral neural activity in the middle-latency evoked field of primary auditory cortex in humans(1). *Hear. Res.* **174**, 19–31. (doi:10.1016/S0378-5955(02)00614-7)

Scott, S. K. & Johnsrude, I. 2003 The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107. (doi:10.1016/S0166-2236(02)00037-1)

Seltzer, B. & Pandya, D. N. 1978 Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* **149**, 1–24. (doi:10.1016/0006-8993(78)90584-X)

Seltzer, B. & Pandya, D. N. 1989*a* Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* **281**, 97–113. (doi:10.1002/cne.902810108)

Seltzer, B. & Pandya, D. N. 1989*b* Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* **290**, 451–471. (doi:10.1002/cne.902900402)

Seltzer, B. & Pandya, D. N. 1991 Post-rolandic cortical projections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* **312**, 625–640. (doi:10.1002/cne.903120412)

Sheffert, S. M., Pisoni, D., Fellowes, J. & Remez, R. 2002 Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 1447–1469. (doi:10.1037/0096-1523.28.6.1447)

Terhardt, E. 1974 Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* **55**, 1061–1069. (doi:10.1121/1.1914648)

Thierry, G., Giraud, A. L. & Price, C. 2003 Hemispheric dissociation in access to the human semantic system. *Neuron* **38**, 499–506. (doi:10.1016/S0896-6273(03)00199-5)

Unoki, M., Irino, T., Glasberg, B., Moore, B. C. J. & Patterson, R. D. 2006 Comparison of the roex and gammachirp filters as representations of the auditory filter. *J. Acoust. Soc. Am.* **120**, 1474–1492. (doi:10.1121/1.2228539)

Uppenkamp, S., Johnsrude, I., Patterson, R. D., Norris, D. & Marslen-Wilson, W. 2006 Locating the initial stages of speech-sound processing in human temporal cortex. *NeuroImage* **31**, 1284–1296. (doi:10.1016/j.neuroimage.2006.01.004)

van Dinther, R. & Patterson, R. D. 2006 Perception of acoustic scale and size in musical instrument sounds. *J. Acoust. Soc. Am.* **120**, 2158–2176. (doi:10.1121/1.2338295)

Vlaardingerbroek, M. & den Boer, J. A. 1996. *Magnetic resonance imaging*. New York, NY: Springer.

von Kriegstein, K. & Giraud, A. L. 2004 Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage* **22**, 948–955. (doi:10.1016/j.neuroimage.2004.02.020)

von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D. & Griffiths, T. D. 2006 Processing the acoustic effect of size in speech sounds. *NeuroImage* **32**, 368–375. (doi:10.1016/j.neuroimage.2006.02.045)

Wallace, M. N., Johnston, P. W. & Palmer, A. R. 2002 Histochemical identification of cortical areas in the auditory region of the human brain. *Exp. Brain Res.* **143**, 499–508. (doi:10.1007/s00221-002-1014-z)

Watkins, K. & Paus, T. 2004 Modulation of motor excitability during speech perception: the role of Broca's area. *J. Cogn. Neurosci.* **16**, 978–987. (doi:10.1162/0898929041502616)

Watkins, K. E., Strafella, A. P. & Paus, T. 2003 Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* **41**, 989–994. (doi:10.1016/S0028-3932(02)00316-0)

Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. 2004 Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* **7**, 701–702. (doi:10.1038/nn1263)

Yeterian, E. H. & Pandya, D. N. 1998 Corticostriatal connections of the superior temporal region in rhesus monkeys. *J. Comp. Neurol.* **399**, 384–402. (doi:10.1002/(SICI)1096-9861(19980928)399:3<384::AID-CNE7>3.0.CO;2-X)

Yost, W. A., Patterson, R. & Sheft, S. 1996 A time domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* **99**, 1066–1078. (doi:10.1121/1.414593)