

- 27 Michelsen, A. *et al.* (1994) Physics of directional hearing in the cricket *gryllus bimaculatus*. *J. Comp. Physiol. Ser. A* 175, 153–162
- 28 Webb, B. (1996) A robot cricket. *Sci. Am.* 275, 94–99
- 29 Pfeifer, R. (2000) On the relation among morphology and materials in adaptive behavior. In *From Animals to Animats 6: Proc. 6th Int. Conf. Simulation Adapt. Behav.* (Meyer, J.A. *et al.*, eds), MIT Press
- 30 Harvey, I. *et al.* (1994) Seeing the light: artificial evolution, real vision. In *From Animals to Animats 3: Proc. 3rd Int. Conf. Simulation Adapt. Behav.* (Cliff, D. *et al.*, eds), MIT Press
- 31 Lichtensteiger, L. and Eggenberger, P. (1999) Evolving the morphology of a compound eye on a robot. In *Proc. 3rd Eur. Workshop Adv. Mobile Robots (Eurobot '99)*, IEEE Press
- 32 Lund, H.H. *et al.* (1997) Evolving robot morphology. In *Proc. 4th Int. Conf. Evol. Comput.*, IEEE Press
- 33 Funes, P. and Pollack, J. (1998) Evolutionary body building: adaptive physical designs for robots. *Artif. Life* 4, 337–357
- 34 Lipson, H. and Pollack, J.B. (2000) Automatic design and manufacture of robotic lifeforms. *Nature* 406, 974–978
- 35 Nolfi, S. and Floreano, D. (1999) Learning and evolution. *Auton. Robots* 7, 89–113
- 36 Todd, P.M. and Miller, G.F. (1991) Exploring adaptive agency II: simulating the evolution of associative learning. In *From Animals to Animats: Proc. 1st Int. Conf. Simulation Adapt. Behav.* (Meyer, J.A. and Wilson, S.W., eds), MIT Press
- 37 Floreano, D. and Nolfi, S. (1997) Adaptive behavior in competing co-evolving species. In *Proc. 4th Eur. Conf. Artif. Life*. (Husband, P. and Harvey, I., eds), MIT Press
- 38 Urzelai, J. and Floreano, D. Evolution of adaptive synapses: robots with fast adaptive behavior in new environments. *Evol. Comput.* (in press)
- 39 Hinton, G.E. and Nowlan, S.J. (1987) How learning guides evolution. *Complex Syst.* 1, 495–502
- 40 Nolfi, S. *et al.* (1994) Learning and evolution in neural networks. *Adapt. Behav.* 3, 5–28
- 41 Nolfi, S. (1999) How learning and evolution interact: the case of a learning task which differs from the evolutionary task. *Adapt. Behav.* 7, 231–236
- 42 Husbands, P. *et al.* (1999) Better living through chemistry: evolving GasNets for robot control. *Connection Sci.* 10 (3–4), 185–210
- 43 Nolfi, S. and Floreano, D. (1998) Co-evolving predator and prey robots: do 'arm races' arise in artificial evolution? *Artif. Life* 4, 311–335

# Structure and function of auditory cortex: music and speech

Robert J. Zatorre, Pascal Belin and Virginia B. Penhune

We examine the evidence that speech and musical sounds exploit different acoustic cues: speech is highly dependent on rapidly changing broadband sounds, whereas tonal patterns tend to be slower, although small and precise changes in frequency are important. We argue that the auditory cortices in the two hemispheres are relatively specialized, such that temporal resolution is better in left auditory cortical areas and spectral resolution is better in right auditory cortical areas. We propose that cortical asymmetries might have developed as a general solution to the need to optimize processing of the acoustic environment in both temporal and frequency domains.

Music and speech represent the most cognitively complex uses of sound by the human species. These two domains share a number of properties, including the fact that they take advantage of modulations of acoustic parameters specifically for information-bearing purposes. Moreover, both music and speech are characterized by their generative nature: that is, complexity is built up by rule-based permutations of a limited number of discrete elements (phonemes or tones) to yield meaningful structures (words or melodies), which in turn are subject to further hierarchical organization resulting in more complex entities (such as sentences or songs) [1]. Apart from these formal considerations, music and speech show other interesting similarities. For example, both show specific and relatively fixed developmental time courses [2], and all known human societies make use of both speech and music, regardless of technological sophistication [3].

The foregoing ideas do not imply that music and speech necessarily share either a similar underlying cognitive or neural representation (for further

discussion see Refs [4,5]), but they do suggest that both might derive from certain functional properties of our auditory nervous system. Put another way, given that all normal humans seem to be capable of relatively sophisticated musical and speech functions in the absence of explicit training, then it follows that these cognitive-behavioral skills are likely to be related to the functional organization of the human auditory nervous system. Just as a bat's ability to use echolocation is related to the unique organization of its auditory system, so the human nervous system may be considered as being organized such that it enables people to readily understand speech and music of the culture in which they are raised. This article explores what this organization might be, with the aim of throwing light upon the neural mechanisms responsible for the low-level perceptual input stage that is relevant for music and speech, and, conversely, taking advantage of speech and music to understand the function of the auditory cortex.

## Speech versus music: different acoustic features?

Before considering the evidence regarding neural specializations, it is useful to consider some relevant acoustic properties of speech and music. Perhaps the most obvious differences between the two relate to the fact that speech is produced by a single 'instrument' – the human voice, whereas music can be produced by practically anything capable of generating sound, including, of course, the voice. However, certain acoustic parameters have been

Robert J. Zatorre\*  
Montreal Neurological  
Institute, 3801 University St.,  
Montreal, Québec,  
Canada H3A 2B4.  
\*e-mail:  
robert.zatorre@mcgill.ca

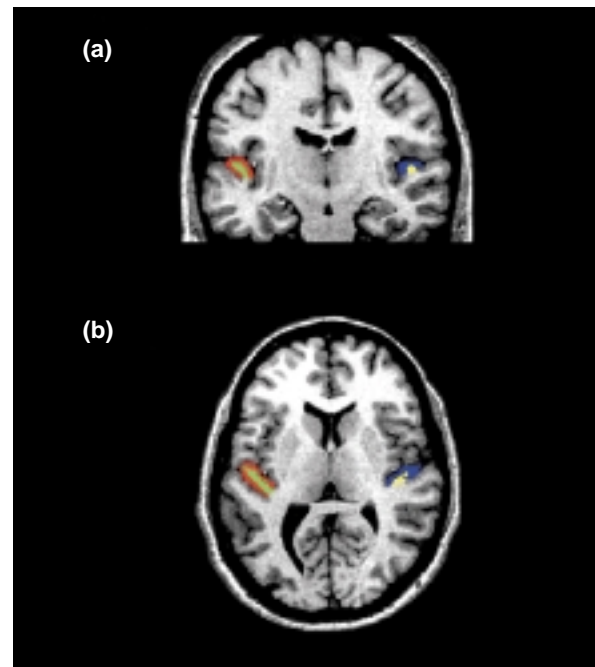
Pascal Belin  
Dept of Psychology,  
Université de Montréal,  
CP 6128 succ Centre-Ville,  
Montreal, Québec,  
Canada H3C 3J7.

Virginia B. Penhune  
Dept of Psychology,  
Concordia University,  
7141 Sherbrooke West,  
Montreal, Québec, Canada  
H4B 1R6.

### Box 1. Anatomical asymmetries in human auditory cortex

Human primary auditory cortex (PAC) is found on the superior surface of the temporal lobe in the region known as Heschl's gyrus (HG). In 1930, von Economo and Horn made the first quantitative measures of HG and found that it was often longer and wider in the left hemisphere than in the right [a]. However, it was not clear whether the extent of PAC differed between the hemispheres. Cytoarchitectonic studies examining the relationship between the location and extent of PAC and HG show that PAC usually covers the most medial two-thirds of the most anterior HG [b,c]. Based on this, Penhune *et al.* used MRI to estimate the location and extent of the region of PAC from the scans of normal subjects [d] (Fig. I). This study revealed a L>R asymmetry, which is the result of a greater volume of white matter underlying HG in the left hemisphere (Fig. II). This asymmetry was proposed to be the result of asymmetries in the cellular organization of these regions.

This hypothesis is supported by a number of findings from cytoarchitectonic studies in the human brain. In PAC, layer III pyramidal cells have been shown to be larger on the left than on the right [e]. Such large cells would be likely to form larger columns, and to send out thicker or more heavily branched axons to other regions of auditory cortex. Supporting this possibility, cell columns in left PAC have been found to be wider and more widely spaced. In addition, each cell column is contacted by more afferents than a similar column on the right [f-h]. A greater number of larger cells, with more heavily myelinated axons and greater interconnectivity could facilitate faster transmission and greater temporal resolution on the left. The complementary structure on



TRENDS in Cognitive Sciences

Fig. I. Coronal (a) and horizontal (b) sections through a magnetic resonance imaging scan of a normal, right-handed volunteer, in which Heschl's gyri are labelled. Gray and white-matter tissue have been segmented and labelled with different colors. Note the greater volume of white matter (green) underlying the left Heschl's gyrus in comparison with the white matter in the right hemisphere (yellow).

identified that are particularly important in transmitting speech, and these tend to emphasize the temporal properties of speech sounds.

Speech contains a large variety of complex sounds of varying temporal grain, including periodic and aperiodic components, noise, frequency and amplitude modulations, and so on, reflecting the many articulatory possibilities of the human vocal musculature. In particular, however, speech scientists have for some time been interested in the acoustic cues that underlie stop consonants, as these seem to be both universal and especially important in understanding speech. Much attention has been focussed on findings that the perception of stop consonants is related to the ability to process temporal differences of the order of tens of milliseconds. Perception of phonetic categories often depends on the relative timing of acoustic events within this range (such as the time between a noise burst and the onset of phonation in a syllable such as /pa/, or the rapidly changing formant transitions characteristic of the stop consonant /d/). This idea is demonstrated most clearly in a study by Shannon *et al.*, who degraded speech by removing most spectral information but retaining temporal variations [6]. The principal result was that relatively good speech comprehension could be obtained with as few as two

spectral channels, indicating that the temporal changes contained within these two noise bands were sufficient to allow the speech decoding mechanism to function adequately.

Additional evidence concerning the importance of temporal cues to speech comes from studies of language impairment. Among the first to make this suggestion was Efron [7], who noted an association between expressive aphasia and temporal judgment deficits; Swisher and Hirsh also found that fluent aphasics with left temporal-lobe damage had the highest temporal ordering thresholds [8] (see also [9]). A similar conclusion regarding the perceptual deficits of patients with pure word deafness was reached by Phillips and Farmer, who commented that the critical problem in these patients relates to a deficit in processing of sounds with temporal content in the range of milliseconds to tens of milliseconds [10]. Tallal and colleagues have reported various strands of evidence suggesting that language acquisition deficits in some children derive from more basic difficulties in processing rapidly changing temporal information [11].

Characterizing the critical acoustic features of music is even more difficult than for speech, because music is far more diverse acoustically, and less research has been carried out on music than on

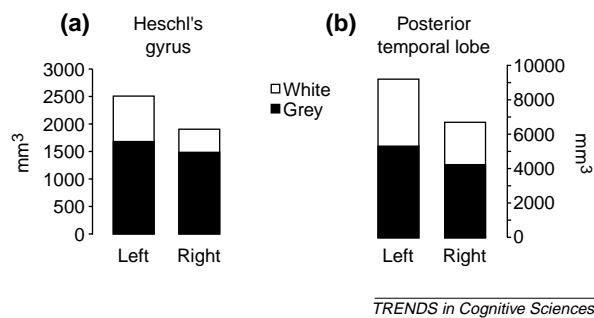


Fig. II. Histograms showing volumes of grey- and white-matter tissue in left and right Heschl's gyrus (a) and posterior temporal lobe (b). Data from the Heschl's gyrus are taken from in-vivo MRI volumetric measures after conversion into standardized stereotaxic space [d]. Data from the posterior temporal lobe are taken from volume measures in post-mortem brain samples [i].

the right could support finer spectral or frequency discrimination.

Studies in post-mortem brain tissue of the adjacent auditory association area, the planum temporale have also shown a L>R asymmetry in white-matter volume, with a similar ratio of grey/white difference as observed in the MRI study of Heschl's gyrus [i] (Fig. IIa). Furthermore, electron microscopy showed that axons in the left planum region were more heavily myelinated than those in the right [i]. Galuske *et al.* showed a L>R asymmetry in the spacing of patches of intrinsic connections in the planum region and hypothesized that this pattern of clustering might allow a more fine-grained analysis of features required for

processing speech [j]. Taken together, this differential organization of left and right auditory regions is compatible with the possibility that left auditory cortical areas have a higher degree of temporal sensitivity, optimal for speech discrimination, whereas corresponding areas on the right have a greater spectral sensitivity, optimal for frequency processing.

#### References

- a von Economo, C. and Horn, L. (1930) Über Windungsrelief, Maße und Rindenarchitektonik der Supratemporalfläche, ihre individuellen und ihre Seitenunterschiede. *Z. Neurol. Psychiat.* 130, 678–757
- b Morosan, P. *et al.* (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13, 684–701
- c Rademacher, J. *et al.* (1993) Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping and neurobiology. *Cereb. Cortex* 3, 313–329
- d Penhune, V.B. *et al.* (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* 6, 661–672
- e Hutsler, J. and Gazzaniga, M. (1996) Acetylcholinesterase staining in human auditory and language cortices: regional variation of structural features. *Cereb. Cortex* 6, 260–270
- f Seldon, H. (1981) Structure of human auditory cortex: I. Cytoarchitectonics and dendritic distributions. *Brain Res.* 229, 277–294
- g Seldon, H. (1981) Structure of human auditory cortex: II. Axon distributions and morphological correlates of speech perception. *Brain Res.* 229, 295–310
- h Seldon, H. (1982) Structure of human auditory cortex: III. Statistical analysis of dendritic trees. *Brain Res.* 249, 211–221
- i Anderson, B. *et al.* (1999) Anatomic asymmetries of the posterior superior temporal lobes: a postmortem study. *Neuropsychiatr. Neuropsychol. Behav. Neurol.* 12, 247–254
- j Galuske, R. *et al.* (2000) Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science* 289, 1946–1949

speech. Given the heterogeneity of sounds relevant for music, it is practical to focus on one specific aspect, which we will refer to as tonal pitch. Pitch variations are a critical component of all musical systems, and it is through modulations of pitch that structures such as melodies are created. We therefore take this component of music as a highly relevant feature to analyze, keeping in mind that it is but one aspect of a much more complex whole.

The nature of pitch processing in music is not fully understood. For our purposes, however, it is sufficient to consider that in melodic patterns, specific pitch-interval ratios are critical, and are fairly accurately encoded in memory [12]: deviations from intended pitch-interval ratio values are readily detected by listeners [13]. Moreover, the pitch variations important for most types of musical processing are relatively small. For example, melodies typically use pitch changes of the order of 1/12 to 1/6 of an octave, and the subdivisions of the octave across cultures tend to be in this range [14]. By contrast, pitch variations used for intonation contours in speech tend to be greater than half an octave, and specific interval relationships are not critical [15].

Pitch changes in music tend to occur on a time scale an order of magnitude slower than those of consonants in speech. Rapidly changing

components of tones do contribute significantly to the microstructure of melodies, influencing such parameters as timbre [16], and phrasing [17], but musical notes are typically much longer in duration than the range characteristic of consonant phonemes [18], and melodies with note durations shorter than about 160 ms are very difficult to identify [19].

#### Speech versus music: different neural substrates?

The contribution of specific neural systems to speech or tonal processing, and the degree to which they might be functionally lateralized, depends to a great extent on the specific nature of the process in question. A common difficulty in the literature on this topic is that the nature of the processes under study is not clearly specified. Nonetheless, it is possible to draw certain conclusions from reviewing the literature. With respect to speech processing, a vast body of data indicates that certain aspects of speech decoding depend critically on left auditory cortical regions. Current evidence suggests that there are functional hierarchies, such that early stages of processing depend on core areas bilaterally, with 'higher' word-recognition mechanisms being associated with more anterior and ventral auditory regions in the left hemisphere [20,21].

The nature of the neural code that underlies pitch perception is currently a matter of debate. There is evidence that both the fine temporal structure of the waveform and its spectral transform are coded by auditory neurons. For our purposes, however, we will concentrate not on the nature of single-unit coding, but rather on the contribution of certain cortical regions to higher-levels representations. There are two main sources of evidence: studies with brain-damaged patients and functional imaging experiments in normal subjects. Many experimental studies have explored the abilities of brain-damaged patients to perceive and discriminate melodic patterns. Although the types of disorders observed vary enormously, it is possible to demonstrate a nearly complete dissociation between intact speech function and highly disturbed tonal processing abilities [22]. For example, Peretz and colleagues have shown that patients with bilateral damage to the temporal cortices who display amusia, or a complete inability to perform even the simplest of tonal processing tasks (such as recognizing a familiar tune), nevertheless show no changes in any speech-related ability [23,24].

Despite the fact that lesions producing amusia are often bilateral, evidence from many systematic studies of the behavioral effects of such lesions suggests a relative hemispheric asymmetry. This asymmetry favors the right temporal cortex in tasks that involve perception of pitch patterns or spectral processing [25–34]. However, most of these studies concur that left temporal-lobe damage can also result in disturbances on certain tasks. Interestingly, when musical tasks involve specifically variations in temporal microstructure (such as detection of slight changes in temporal synchrony or duration), performance is impaired with left but not right temporal-lobe damage [35,36].

With certain low-level pitch-processing tasks a specific pattern of results has implicated the right auditory cortex. We showed previously that only lesions encroaching into the right Heschl's gyrus – which contains primary auditory cortex (see Box 1) – resulted in perceptual deficits in perceiving the pitch of a spectrally complex tone with no energy at the fundamental [37]. This finding is supported by recent MEG data [38]. Damage in the right auditory cortex also leads to an increased threshold in determining the *direction* of pitch change between two notes, but not in simple discrimination of those same pitches [39]. Notably, however, the patients in this study were not unable to complete the discrimination task, but rather showed a fourfold increase in their perceptual threshold; hence the lesion did not destroy the ability to make the discrimination *per se*, but rather produced difficulties with fine-grained pitch resolution.

More broadly, relative specialization within right auditory regions for tonal processing is supported by functional imaging data from a wide variety of perceptual tasks, including pitch judgments within

melodies [40], speech syllables [41], or tones [42]; maintenance of pitch while singing [43]; imagery for tunes [44]; discrimination of pitch and duration in short patterns [45]; reproduction of tonal rhythmic patterns [46]; timbre judgments in dichotic stimuli [47]; and detection of deviant chords [48]. Some studies focusing on temporal fine structure have not identified any asymmetries, however [49]. With the heterogeneity of tasks, and therefore the variety of cognitive mechanisms involved in these diverse studies, many different cortical and subcortical regions in both hemispheres have been implicated. However, the overall pattern that emerges is similar to the hierarchical organization seen in speech studies, in that core cortical areas are primarily involved in specific low-level processes related to extraction and ordering of pitch information, whereas 'belt' and 'parabelt' regions tend to be important for processing patterns of tones. Finally, because organizing sounds in time makes important demands on working memory function, interactions with frontal-lobe regions, especially on the right, are frequently observed [22,28,40,41,50].

#### Speech and music: a unifying framework?

If one accepts the evidence of relative hemispheric asymmetries in processing of tonal pitch and speech sounds, then the main question that arises is why such asymmetries exist, and how they have come about. A possible answer is afforded by the consideration of the different acoustical cues that are relevant for each type of signal, as outlined above. Our hypothesis is that functional asymmetries stem from the fundamentally different computational demands made by environmental signals containing information in the time domain or in the spectral domain.

In particular, we propose that there is a relative trade-off between the specializations required for each type of processing, such that enhancing resolution in one domain comes at the expense of the other (see Box 2). A strict reciprocal relationship might be expected only in a linear system, whereas the auditory nervous system is highly non-linear and distributed. However, the auditory system must often deal with less than optimal signal-to-noise conditions, and such adverse circumstances might impose constraints on the optimization of signal processing. According to this view, hemispheric asymmetries in auditory processing arise as a solution to the inherently incompatible requirement that processing of both temporal and spectral information be optimized. Rather than sacrifice one type of processing in favor of the other, therefore, it would be advantageous to develop two parallel and complementary systems – one in each hemisphere – specialized for rapid temporal processing or for fine spectral processing respectively.

Neurophysiological recordings in macaque monkeys have shown that auditory cortical neurons often have complex response profiles that incorporate sensitivity to both spectral and



### Box 2. The acoustic uncertainty principle

One of the most intriguing aspects of the quantum nature of matter is expressed by Heisenberg's 'Uncertainty Principle', which states that it is impossible to make a precise simultaneous measurement of both the position of a particle and its momentum. The mathematical formulation of this principle is:

$$\Delta p_x \cdot \Delta x \geq h/4\pi$$

where  $\Delta x$  is the uncertainty in the position of a particle in the  $x$  direction,  $\Delta p_x$  the uncertainty in its momentum in the  $x$  direction, and  $h$  is Planck's constant ( $6.626 \times 10^{-34}$  J s). In other words, when uncertainties are small enough, there is a trade-off between the precision that a simultaneous measurement can achieve in the two dimensions: the more precise the measure of position, the less precise the measure of momentum, and vice versa.

There exists an analogous principle in acoustics, reflecting an essential indefiniteness in the distribution of energy over time and frequency. The 'Acoustic Uncertainty Principle', explicitly articulated by Joos, states that one cannot make a precise simultaneous measurement of an auditory event in both the time and frequency domains [a]. It can be formulated as follows:

$$\Delta t \cdot \Delta f \geq c$$

where  $\Delta t$  is the uncertainty along the time dimension,  $\Delta f$  the uncertainty along the frequency dimension, and  $c$  is a constant (on the order of 0.5 Hz). As in the corresponding principle in quantum physics, it implies that there is a trade-off between the precision that can be achieved in the time and frequency domains when measuring an acoustic event: at the upper limit of resolution, increasing temporal precision can only be achieved at the expense of frequency precision, and vice versa.

temporal features of sounds (e.g. Refs [51–53]). Eggermont noted that in cat auditory cortex there was a correlation between bandwidth and limiting rate, such that neurons with greater bandwidth tended to have greater sensitivity to temporal changes and vice versa, commenting that this finding 'conforms to the intuitive notion that broader frequency tuning should allow better temporal responses' (Ref. [54], p. 2762; see also Box 3).

Of more direct relevance are data from Liégeois-Chauvel and colleagues, who recorded electrical potentials from the human auditory cortex using implanted electrodes [55]. They observed that responses from left, but not right, Heschl's gyrus distinguished brief temporal differences both in speech and non-speech sounds. More recently, these authors presented data that indicates a sharper tuning to frequency in the right auditory cortex than

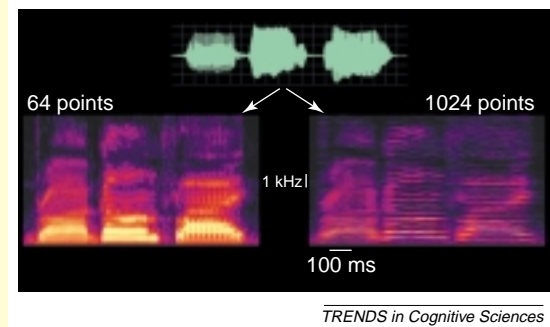


Fig. 1. Spectral-temporal trade-off in spectrograms. The waveform of an excerpt of sung speech (upper panel) is subjected to spectrographic analyses using a large window with 1024 points (right), or a small window with 64 points (left) in the Fourier transform. Individual harmonics are seen as horizontal stripes only in the right panel (good frequency resolution), whereas individual glottal pulses are seen as vertical stripes only in the left panel (good temporal resolution).

This fundamental limitation is particularly evident when acoustic energy is decomposed in time and frequency when computing a sound spectrogram (Fig. 1). A critical parameter is the size of the window used in the Fourier transforms, which corresponds both to the duration of signal analyzed and to the number of different frequency channels used in the analysis. Using a large window yields a high frequency resolution because of the large number of independent frequency channels, but a poor temporal resolution because different brief events could be grouped together in the long window (Fig. 1, right). Conversely, using a small window for the Fourier transform of short duration yields good temporal precision, but poor frequency resolution owing to the small number of independent frequency channels used (Fig. 1, left).

#### Reference

- a Joos, M. (1948) Acoustic phonetics. *Lang. Monogr.* 23, 1–137

in the left [56] (Fig. 1). These data are therefore consistent with the idea that left auditory cortex responses are optimal for a range of temporal responses relevant for speech distinctions, but are not highly selective for frequency, whereas right auditory responses have a greater spectral resolution but are less sensitive to brief temporal events. Poeppel has recently reached a similar conclusion based on MEG data [57]; he suggests that different temporal integration windows exist – the left auditory cortex has a short window and is sensitive to faster events (20–50 ms), whereas the right cortex is sensitive at a slower range (150–250 ms).

These physiological data are consistent with much of the lesion literature described above, and particularly that of Robin *et al.*, who noted a double dissociation between left-hemisphere lesioned patients, who were impaired at tasks requiring high

### Box 3. Auditory processing asymmetries in humans and other species.

A recurring controversy in cognitive neuroscience is the extent to which processes believed to be operative in humans are also reflected in other species. In the case of auditory cortical organization, the evidence is incomplete, and therefore any conclusions are subject to revision. Nevertheless, it is interesting to consider some reports of auditory cortical specialization which are relevant to the current discussion. Hemispheric asymmetries are only occasionally reported in animal studies, but it is not clear if this is because they are weak or nonexistent, or because they are not often tested for. There is evidence, however, that unilateral left but not right lesions to the temporal cortex in the monkey result in deficits, albeit transient ones, in discriminating vocalizations [a]. Behavioral studies also indicate a right-ear preference for monkey calls, which tends to disappear if the sounds are temporally stretched [b].

In addition, several studies have suggested that there might be asymmetries in the processing of rapid acoustic cues in rodents. For example, rate-dependent left-hemisphere lateralization effects for rapidly changing tone sequences in the rat have been reported [c], although subsequent lesion studies reported bilateral effects [d]. Asymmetries have also been reported at subcortical levels: evoked responses to speech-like stimuli measured from the medial geniculate in the guinea pig revealed greater left than right responses [e]. Conversely, there is also some limited evidence that right auditory cortex may be important for processing frequency information: in the Mongolian gerbil, discrimination of rising versus falling tone glides is affected only by right and not left auditory cortex lesions [f] (Fig. 1).

These relatively few observations are far from proving that identical mechanisms exist for cortical auditory processing between humans and other animals. Nor do these findings necessarily mean that the processes in question emerged from a common ancestor, since convergent evolution may have led to any similarities. It also appears that the most complex generative and hierarchical aspects of human auditory processing that are most characteristic of speech and tonal pitch function are largely absent in other species. Nonetheless, enough analogies would appear to exist that further investigation would be warranted. If the specializations seen in other species are indeed eventually shown to be related to those present in humans, it would support the idea that they may

have emerged as general solutions to the need to optimize both temporal and spectral processing, rather than being directly related to higher-order linguistic or music-related processes.

#### References

- a Heffner, H. and Heffner, R. (1984) Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 226, 75–76
- b Hauser, M. (2000) The sound and the fury: primate vocalizations as reflections of emotion and thought. In *The Origins of Music* (Wallin, N. *et al.*, eds), pp. 77–102, MIT Press
- c Fitch, R. *et al.* (1993) Functional lateralization for auditory temporal processing in male and female rats. *Behav. Neurosci.* 107, 844–850
- d Fitch, R. *et al.* (1994) Induced microgyria and auditory temporal processing in rats: a model for language impairment? *Cereb. Cortex* 4, 260–270
- e King, C. *et al.* (1999) Thalamic asymmetry is related to acoustic signal complexity. *Neurosci. Lett.* 267, 89–92
- f Wetzel, W. *et al.* (1998) Right auditory cortex lesion in Mongolian gerbils impairs discrimination of rising and falling frequency-modulated tones. *Neurosci. Lett.* 252, 115–118

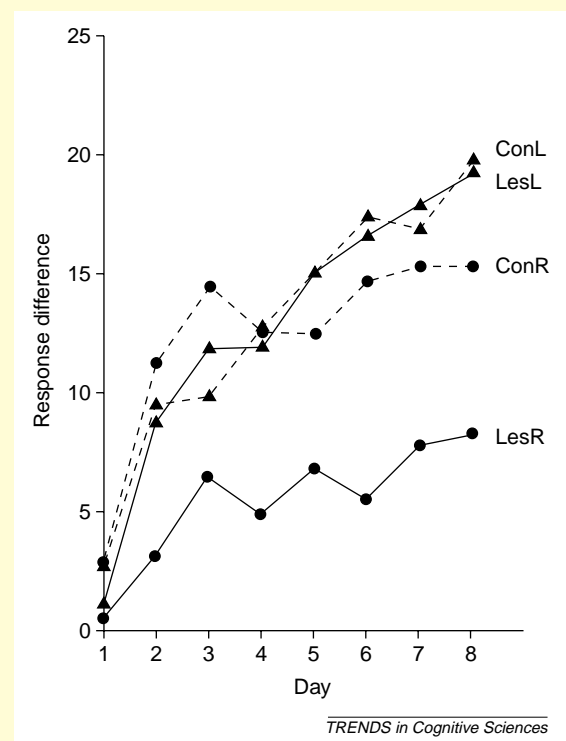


Fig. 1. Effect of unilateral auditory cortex lesions in the Mongolian gerbil on discrimination of frequency-modulated tones. LesL and LesR: animals with left or right lesions; ConL and ConR: animals with control lesions. Mean performance (correct conditioned responses minus false alarms) is plotted as a function of days of training. Adapted with permission from Ref. [f].

temporal but not spectral resolution, and right-hemisphere lesioned patients, who showed the opposite pattern [33]. Enhanced temporal resolution in left auditory cortex was also shown in a PET study

by Belin *et al.*, in which the rate of formant transitions in a pseudospeech syllable was either 40 ms or 200 ms [58]. Cerebral blood flow increases in left auditory regions were found for both types of

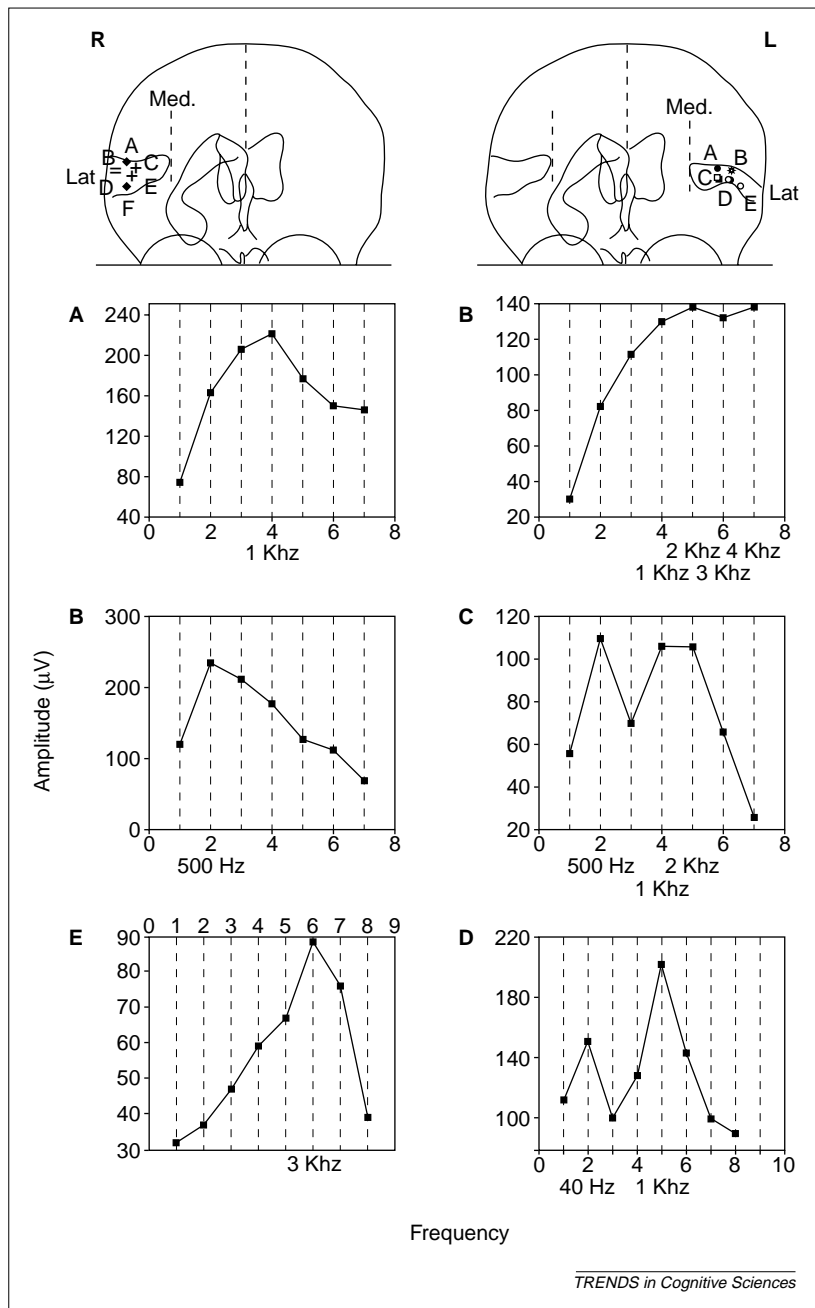


Fig. 1. Results of stimulation with pure tones from left and right Heschl's gyrus in human patients during depth-electrode recordings. Graphs show the amplitude variation of the 80-ms component of the evoked response as a function of frequency. Note that responses from the left auditory cortex (right side of figure, individual records shown in B,C,D) more often had multiple peaks or showed broad tuning, whereas responses from the right auditory cortex (left side of figure, A, B, E) tended to show sharper, single-peaked responses. Reproduced with permission from Ref. [56].

stimuli, indicating a capacity for processing spectral change over a wide range of durations; by contrast, activity in regions of right auditory cortex was observed only to the slower rate and not to the faster rate. Thus, the right auditory cortex seems to respond less well to fast formant transitions, whereas the left is better able to track rapidly changing acoustic information, which would be relevant for speech processing, even though the stimuli in this experiment were not perceived as speech.

### Spectral and temporal sensitivity in auditory cortex

A recent functional imaging study in our laboratory set out to test the interaction between spectral and temporal processing more directly, using a parametric approach [59]. Two stimulus parameters, rate of frequency modulation, and spectral distribution of elements within the pattern, were varied independently in the two conditions of the experiment (stimuli may be heard at: [www.zlab.mcgill.ca](http://www.zlab.mcgill.ca)). The prediction was that increasing the rate of temporal change would preferentially recruit left auditory cortical areas, whereas increasing the number of spectral elements would engage right auditory cortical regions.

We found that distinct areas of auditory cortex in each hemisphere responded to each parameter (Fig. 2a); but most pertinent for the present discussion, cerebral blood flow in the anterior auditory region on the right side showed a greater response to increasing spectral than temporal variation, whereas a symmetrical area on the left showed the reverse pattern. A third area, located in the right superior temporal sulcus, also showed a significant response to the spectral parameter, but showed no change to the temporal parameter (Fig. 2b). Finally, a direct comparison of the temporal and spectral conditions revealed greater left cortical activity for the combined temporal scans, but greater right activity for the combined spectral scans (Fig. 2c).

The finding that distinct areas in both hemispheres respond to the different parameters fits with recent views about the hierarchical arrangement of primate auditory cortex [60]. The observation that a region in the upper bank of the right superior temporal sulcus responds to increasing spectral complexity is concordant with recent fMRI findings that this area is sensitive to the acoustic shape of human vocal sounds [61], because such processing would require sensitivity to subtle changes in spectral energy distribution. More generally, our findings are in agreement with the proposal that these regions form part of a ventral stream specialized for object-feature processing, in which spectral information would play an important role [62].

The fact that cortical areas in both hemispheres respond to both temporal and spectral change is consistent with findings that disorders such as pure word deafness are usually associated with bilateral damage [10], and with imaging studies, which usually show bilateral activation [21]. Above and beyond the activity being bilateral, however, clear hemispheric differences emerged in the predicted direction. Thus, the data support the hypothesis that there is a *relative* specialization of the auditory cortices in the two hemispheres, with a reciprocal relationship between temporal and spectral resolution. We speculate that maximizing the processing of rapidly changing information (high temporal resolution) might set an upper limit on

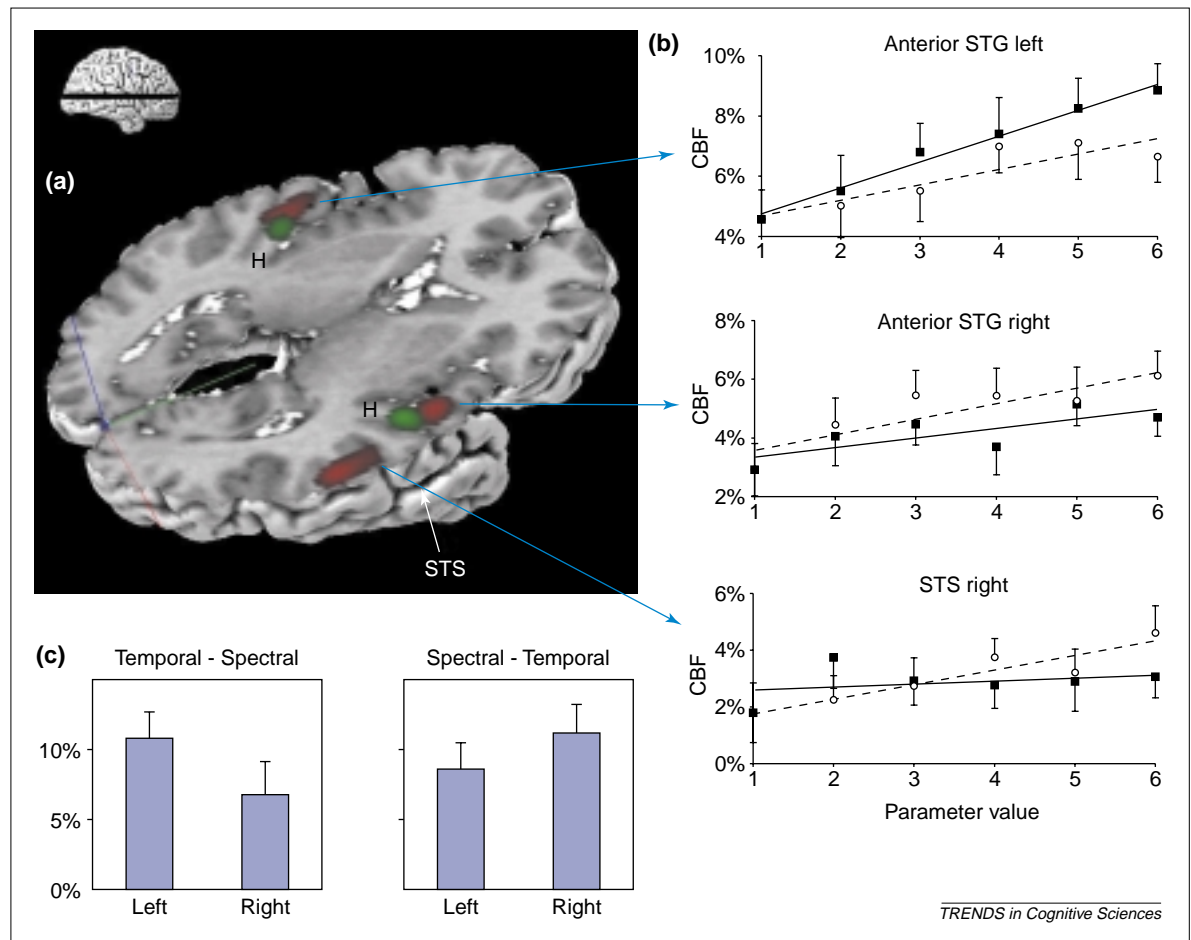


Fig. 2. (a) 3-D rendering of the functional activation pattern in our spectral-temporal study. The brain is viewed from the right side, with the level of the horizontal section indicated in the inset. Green areas, located bilaterally within Heschl's gyrus (H) correspond to regions showing significant covariation of CBF with increasing rate of temporal change, red areas (located in anterior superior temporal cortex, and in right superior temporal sulcus, STS) correspond to regions whose CBF increased as a function of changes in the spectral parameter. (b) Blood flow changes in the three cortical regions that showed increases to the spectral input parameter (open symbols, spectral parameter; filled symbols, temporal parameter). Symbols indicate average cerebral blood flow (CBF)  $\pm$  one standard error of the mean; lines are drawn through the least-squares linear solution corresponding to these points. (c) Histograms showing percentage blood flow difference when all temporal and spectral conditions were compared with one another. The left graph shows greater activity in the left than the right Heschl's gyrus (green areas in a) in response to the temporal stimuli as compared with the spectral stimuli; the right graph shows greater activity in the right than left anterior superior temporal cortex (red areas in a) to the spectral as compared with the temporal stimuli. Adapted from Ref. [59].

frequency resolution; this could explain why the changes in cerebral blood flow observed within the left anterior STG area were greater to the temporal parameter than to the spectral parameter. Conversely, higher frequency resolution might entail a relative decrement in temporal processing, hence the inverse pattern seen in the right anterior STG and STS regions. This model does raise the question of how such functional differences might be instantiated in the brain – but at least a partial answer is afforded by studies of the

anatomical differences between left and right human auditory cortex (Box 1).

### Conclusion

The present model suggests that relatively subtle quantitative differences in neural response properties early in the cortical processing stream can lead to qualitatively distinct functional roles for higher-order processes. Higher-order specializations, involving brain areas beyond the auditory cortices (see Box 4), might arise as a consequence of the organization of intra-hemispheric local circuits [63]. Thus, the predominant role of the left hemisphere in many complex linguistic functions might have arisen from a slight initial advantage in decoding speech sounds. The important role of the right hemisphere in aspects of musical perception – particularly those involving tonal pitch processing – might then have been in some sense a consequence of, and is complementary to, this specialization of language. An advantageous feature of the hypothesis presented above is that it offers a unifying framework for understanding some of the functional characteristics of the auditory nervous system that are relevant for processing speech and tonal patterns. The complementary specialization of the auditory cortices in the two hemispheres would thus be seen as arising from a single underlying principle rather than being unrelated phenomena.



#### Box 4. Speech versus music: low-level features or domain specificity?

The data and model presented in this article clearly emphasize the importance of low-level cues in distinguishing between speech and tonal pitch processing. In particular, the proposed differences between spectral and temporal resolution in the left and right auditory cortices are predicated on the idea that speech and music exploit different ranges of temporal and spectral processing, and that hemispheric differences emerge as a consequence of differences in processing these cues, and not from domain-specific functions. This view is supported by findings that stimuli that are not at all speech-like tend to engage or depend on left auditory cortices if they contain certain temporal features (see main text).

Arguing against this view are data that point to domain specificity in both speech and music processing. A good example of this phenomenon is afforded by neuroimaging studies of linguistic pitch processing in tonal languages, such as Thai or Chinese, which have demonstrated that left-hemisphere structures are recruited for processing pitch contours specifically in speakers of such languages, whereas non-speakers process the identical stimuli via right-hemisphere mechanisms [a–c]. A similar argument has been made based on electrical and magnetic recordings indicating language-specific differences in processing vowels in the left temporal cortex [d]. Perhaps the most dramatic evidence that left-hemisphere structures mediate abstract properties of language rather than certain acoustic features comes from studies of left-hemisphere involvement in visual-sign processing in the deaf [e,f]. More generally, it has been argued on the basis of a variety of empirical and theoretical considerations that speech processing depends on a specific and dedicated neural circuitry not used for other types of processes [g], and that the same may be true for music [h].

A resolution to these issues is yet to be reached, but perhaps the two approaches need not be mutually exclusive (as is sometimes thought). It might be possible for functional

specialization to emerge at multiple levels and interact in complex ways. In particular, we suggest that low-level specializations evolved not as a function of domain-specific processes, but as a general solution to the computational problem posed by the need for precision of both spectral and temporal processing. Subsequent stages of processing might initially have developed from these low-level specializations, as small differences in processing are likely to be amplified at each node when one goes to higher levels in the functional architecture. However, this need not imply that these higher-order mechanisms are *only* tied to the nature of the initial processing. Indeed, once the neural mechanisms responsible for abstract high-level aspects of language or music have developed, there may well be multiple routes of access, independently of the nature of the input. In this context, it is noteworthy that many of the above examples of domain-specific processing involve regions outside the auditory cortex, especially frontal areas, which could thus be independent of low-level auditory specializations.

#### References

- a Gandour, J. *et al.* (1998) Pitch processing in the human brain is influenced by language experience. *NeuroReport* 9, 2115–2119
- b Gandour, J. *et al.* (2000) A crosslinguistic PET study of tone perception. *J. Cogn. Neurosci.* 12, 207–222
- c Klein, D. *et al.* (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage* 13, 646–653
- d Näätänen, R. *et al.* (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434
- e Hickok, G. *et al.* (1998) The neural organization of language: evidence from sign language aphasia. *Trends Cogn. Sci.* 2, 129–136
- f Petitto, L. *et al.* (2000) Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13961–13966
- g Liberman, A.M. and Whalen, D.H. (2000) On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196
- h Peretz, I. (2001) Brain specialization for music: new evidence from congenital amusia. *Ann. New York Acad. Sci.* 930, 153–165

#### Questions for future research

- Recent evidence suggests that there are at least two and possibly more cortical processing streams concerned with different aspects of auditory processing (object-related features, spatial features, distribution of spectral energy over time, etc.). How does the proposed spectral/temporal distinction between the auditory cortices in the two hemispheres interact with these cortical pathways?
- What is the nature of the specialization beyond auditory cortex? How do the putative differences in auditory cortex relate to specializations in frontal, parietal or other regions?
- We have emphasized differences between speech and music processing, but there are also similarities. Can these complex abilities be broken down into separate sub-components in order to identify both points of commonality and points of divergence?
- Can anatomical features of auditory cortical organization predict individual patterns of hemispheric specialization for speech? Are these features genetically predetermined or do they change as a function of learning?
- How does auditory cortical functional specialization change as a consequence of damage? Do the features discussed in this article impose constraints on reorganization?
- Mapping of homologies between human and other species has been successful in the visual system. Can a similar approach also reap benefits in the case of the auditory system? Do hemispheric differences found in humans always have parallels in other species, or are there specializations unique to each species?

#### Acknowledgements

This work was supported by funding from the Canadian Institutes of Health Research and by the McDonnell-Pew Cognitive Neuroscience Program.

## References

- 1 Krumhansl, C. (1990) *Cognitive Foundations of Musical Pitch*, Oxford University Press
- 2 Trehub, S.E. (2001) Musical predispositions in infancy. *Ann. New York Acad. Sci.* 930, 1–16
- 3 Brown, D. (1991) *Human Universals*, McGraw-Hill
- 4 Besson, M. and Schön, D. (2001) Comparison between language and music. *Ann. New York Acad. Sci.* 930, 232–258
- 5 Brown, S. (2000) The 'musilanguage' model of music evolution. In *The Origins of Music* (Wallin, N. et al., eds), pp. 271–300, MIT Press
- 6 Shannon, R.V. et al. (1995) Speech recognition with primarily temporal cues. *Science* 270, 303–304
- 7 Efron, R. (1963) Temporal perception, aphasia and déjà vu. *Brain* 86, 403–423
- 8 Swisher, L. and Hirsh, I.J. (1972) Brain damage and the ordering of two temporally successive stimuli. *Neuropsychologia* 10, 137–152
- 9 von Steinbüchel, N. (1998) Temporal ranges of central nervous processing: clinical evidence. *Exp. Brain Res.* 123, 220–233
- 10 Phillips, D.P. and Farmer, M.E. (1990) Acquired word deafness and the temporal grain of sound representation in the primary auditory cortex. *Behav. Brain Res.* 40, 84–90
- 11 Tallal, P. et al. (1993) Neurobiological basis of speech: A case for the preeminence of temporal processing. *Ann. New York Acad. Sci.* 682, 27–47
- 12 Attneave, F. and Olson, R.K. (1971) Pitch as a medium: a new approach to psychophysical scaling. *Am. J. Psychol.* 84, 147–166
- 13 Warrier, C. and Zatorre, R.J. Influence of tonal context and timbral variation on perception of pitch. *Percept. Psychophys.* (in press)
- 14 Vos, P. and Troost, J. (1989) Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396
- 15 Patel, A. et al. (1998) Processing prosodic and musical patterns: A neuropsychological investigation. *Brain Lang.* 61, 123–144
- 16 Miller, J. and Carterette, E. (1975) Perceptual space for musical structures. *J. Acoust. Soc. Am.* 58, 711–720
- 17 Repp, B. (1992) Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei'. *J. Acoust. Soc. Am.* 92, 2546–2568
- 18 Fraisse, P. (1974) *Psychologie du rythme*, Presses Universitaires de France
- 19 Warren, R. et al. (1991) Melodic and nonmelodic sequences of tones: effects of duration on perception. *Music Percept.* 8, 277–289
- 20 Binder, J. et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528
- 21 Hickok, G. and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138
- 22 Marin, O.S.M. and Perry, D.W. (1999) Neurological aspects of music perception and performance. In *The Psychology of Music* (Deutsch, D., ed.), pp. 653–724, Academic Press
- 23 Peretz, I. et al. (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117, 1283–1301
- 24 Ayotte, J. et al. (2000) Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain* 123, 1926–1938
- 25 Milner, B.A. (1962) Laterality effects in audition. In *Interhemispheric Relations and Cerebral Dominance* (Mountcastle, V., ed.), pp. 177–195, Johns Hopkins Press
- 26 Divenyi, P. and Robinson, A. (1989) Nonlinguistic auditory capabilities in aphasia. *Brain Lang.* 37, 290–326
- 27 Samson, S. and Zatorre, R.J. (1994) Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia* 32, 231–240
- 28 Zatorre, R.J. and Samson, S. (1991) Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain* 114, 2403–2417
- 29 Zatorre, R.J. (1985) Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia* 23, 31–41
- 30 Zatorre, R.J. and Halpern, A.R. (1993) Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia* 31, 221–232
- 31 Liégeois-Chauvel, C. et al. (1998) Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867
- 32 Sidtis, J.J. and Volpe, B.T. (1988) Selective loss of complex-pitch or speech discrimination after unilateral lesion. *Brain Lang.* 34, 235–245
- 33 Robin, D.A. et al. (1990) Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang.* 39, 539–555
- 34 Penhune, V.B. et al. (1999) The role of auditory cortex in retention of rhythmic patterns in patients with temporal-lobe removals including Heschl's gyrus. *Neuropsychologia* 37, 315–331
- 35 Ehrlich, N. et al. (2001) Processing of rapid auditory information in epileptic patients with left temporal lobe damage. *Neuropsychologia* 39, 525–531
- 36 Samson, S. et al. (2001) Cerebral substrates for musical temporal processes. *Ann. New York Acad. Sci.* 930, 166–178
- 37 Zatorre, R.J. (1988) Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566–572
- 38 Patel, A. and Balaban, E. (2001) Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat. Neurosci.* 4, 839–844
- 39 Johnsrude, I.J. et al. (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155–163
- 40 Zatorre, R.J. et al. (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919
- 41 Zatorre, R.J. et al. (1992) Lateralization of phonetic and pitch processing in speech perception. *Science* 256, 846–849
- 42 Binder, J. et al. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362
- 43 Perry, D.W. et al. (1999) Localization of cerebral activity during simple singing. *NeuroReport* 10, 3979–3984
- 44 Halpern, A.R. and Zatorre, R.J. (1999) When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* 9, 697–704
- 45 Griffiths, T.D. et al. (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport* 10, 3825–3830
- 46 Penhune, V.B. et al. (1998) Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* 10, 752–765
- 47 Huggdahl, K. et al. (1999) Brain activation during dichotic presentations of consonant–vowel and musical instrument stimuli: a 15O-PET study. *Neuropsychologia* 37, 431–440
- 48 Tervaniemi, M. et al. (2000) Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79
- 49 Griffiths, T.D. et al. (1998) Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.* 1, 422–427
- 50 Holcomb, H. et al. (1998) Cerebral blood flow relationships associated with difficult tone recognition task in trained normal volunteers. *Cereb. Cortex* 8, 534–542
- 51 deCharms, R.C. et al. (1998) Optimizing sound features for cortical neurons. *Science* 280, 1439–1443
- 52 Phillips, D.P. (1993) Representation of acoustic events in the primary auditory cortex. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 203–216
- 53 Steinschneider, M. et al. (1995) Physiologic correlates of the voice onset time boundary in primary auditory cortex of the awake monkey: temporal response patterns. *Brain Lang.* 48, 326–340
- 54 Eggermont, J. (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. *J. Neurophysiol.* 80, 2743–2764
- 55 Liégeois-Chauvel, C. et al. (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb. Cortex* 9, 484–496
- 56 Liégeois-Chauvel, C. et al. (2001) Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of the human auditory cortex. *Ann. New York Acad. Sci.* 930, 117–132
- 57 Poeppel, D. Pure word deafness and the bilateral processing of the speech code. *Cogn. Sci.* (in press)
- 58 Belin, P. et al. (1998) Lateralization of speech and auditory temporal processing. *J. Cogn. Neurosci.* 10, 536–540
- 59 Zatorre, R.J. and Belin, P. (2001) Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953
- 60 Kaas, J.H. et al. (1999) Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170
- 61 Belin, P. et al. (2000) Voice-selective areas in human auditory cortex. *Nature* 403, 309–312
- 62 Rauschecker, J.P. and Tian, B. (2000) Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11800–11806
- 63 Ringo, J. et al. (1994) Time is of the essence: a conjecture that hemispheric specialization arises from interhemispheric conduction delay. *Cereb. Cortex* 4, 331–343