

# Neural specializations for speech and pitch: moving beyond the dichotomies

Robert J. Zatorre<sup>1,\*</sup> and Jackson T. Gandour<sup>2</sup>

<sup>1</sup>*Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, H3A 3B4 Quebec, Canada*

<sup>2</sup>*Purdue University, West Lafayette, IN 47907, USA*

The idea that speech processing relies on unique, encapsulated, domain-specific mechanisms has been around for some time. Another well-known idea, often espoused as being in opposition to the first proposal, is that processing of speech sounds entails general-purpose neural mechanisms sensitive to the acoustic features that are present in speech. Here, we suggest that these dichotomous views need not be mutually exclusive. Specifically, there is now extensive evidence that spectral and temporal acoustical properties predict the relative specialization of right and left auditory cortices, and that this is a parsimonious way to account not only for the processing of speech sounds, but also for non-speech sounds such as musical tones. We also point out that there is equally compelling evidence that neural responses elicited by speech sounds can differ depending on more abstract, linguistically relevant properties of a stimulus (such as whether it forms part of one's language or not). Tonal languages provide a particularly valuable window to understand the interplay between these processes. The key to reconciling these phenomena probably lies in understanding the interactions between afferent pathways that carry stimulus information, with top-down processing mechanisms that modulate these processes. Although we are still far from the point of having a complete picture, we argue that moving forward will require us to abandon the dichotomy argument in favour of a more integrated approach.

**Keywords:** hemispheric specialization; functional neuroimaging; tone languages

## 1. INTRODUCTION

The power of speech is such that it is often considered nearly synonymous with being human. It is no wonder, then, that it has been the focus of important theoretical and empirical science for well over a century. In particular, a great deal of effort has been devoted to understanding how the human brain allows speech functions to emerge. Several distinct intellectual trends can be discerned in this field of research, of which two particularly salient ones will be discussed in this paper. One important idea proposes that speech perception (and production) depends on specialized mechanisms that are dedicated exclusively to speech processing. A contrasting idea stipulates that speech sounds are processed by the same neural systems that are also responsible for other auditory functions.

These two divergent ideas will be referred to here as the domain-specific and the cue-specific models, respectively. Although they are most often cast as mutually exclusive, it is our belief, and the premise of this piece, that some predictions derived from each of these views enjoy considerable empirical support, and

hence must be reconciled. We shall therefore attempt to propose a few ideas in this regard, after first reviewing the evidence that has been adduced in favour of each model. In particular, we wish to review the evidence pertaining to patterns of cerebral hemispheric specialization within auditory cortices, and their relation to the processing of speech signals and non-speech signals, with particular emphasis on pitch information. We constrain the discussion in this way in order to focus on a particularly important aspect of the overall problem, and also one that has generated considerable empirical data in recent years; thus, it serves especially well to illustrate the general question. Also, owing to the development of functional neuroimaging over the past few years, predictions derived from these two models have now been tested to a much greater extent than heretofore feasible. We therefore emphasize neuroimaging studies insofar as they have shed light on the two models, but also mention other sources of evidence when pertinent.

First, let us consider some of the origins of the domain-specificity model. Much of the impetus for this idea came from the behavioural research carried out at the Haskins laboratories by Alvin Liberman and his colleagues in the 1950s and 1960s (for reviews, see Liberman & Mattingly (1985) and Liberman & Whalen (2000)). These investigators took advantage of then newly-developed techniques to visualize speech sounds (the spectrograph), and were struck by the observation that the acoustics of speech sounds did not map in a one-to-one fashion to the perceived

\* Author for correspondence (robert.zatorre@mcgill.ca).

We wish to dedicate this paper to the memory of Peter D. Eimas (1934–2005), a devoted investigator and inspiring teacher who made significant contributions to the field of speech research, and whose thoughts influenced many of the ideas in the present text.

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

phonemes. These findings led to the development of the motor theory of speech perception, which proposes that speech sounds are deciphered not by virtue of their acoustical structure, but rather by reference to the way that they are articulated. The lack of invariance in the signal was explained by proposing that invariance was instead present in the articulatory gestures associated with a given phoneme, and that it was these representations which were accessed in perception. More generally, this model proposed that speech bypassed the normal pathway for analysis of sound, and was processed in a dedicated system exclusive to speech. This view therefore predicts that specialized left-hemisphere lateralized pathways exist in the brain which are unique to speech. A corollary of this view is that low-level acoustical features are not relevant for predicting hemispheric specialization, which is seen instead to emerge only from abstract, linguistic properties of the stimulus. Thus, a strong form of this model would predict, for instance, that certain left auditory cortical regions are uniquely specific to speech and would not be engaged by non-speech signals. An additional prediction is that the linguistic status of a stimulus will change the pattern of neural response (e.g. a stimulus that is perceived as being speech or not under different circumstances, or by different persons, would be expected to result in different neural responses).

An alternative approach to this domain-specific model proposes that general mechanisms of auditory neural processing are sufficient to explain speech perception (for a recent review, see Diehl *et al.* (2004)). Several different approaches are subsumed within what we refer to as the cue-specific class of models, but all of them would argue that speech-unique mechanisms are unnecessary and therefore unparsimonious. In terms of the focus of the present discussion, a corollary of this point of view is that low-level acoustical features of a stimulus can determine the patterns of hemispheric specialization that may be observed; hence, it would predict that certain features of non-speech sounds should reliably recruit the left auditory cortex, and that there would be an overlap between speech- and non-speech-driven neural responses under many circumstances. That is, to the extent that certain left-hemisphere auditory cortical regions are involved with the analysis of speech, this is explained on the basis that speech sounds happen to have particular acoustical properties, and it is the nature of the processing elicited by such properties that is at issue, not the linguistic function to which such properties may be related.

The problem that faces us today is that both models make at least some predictions which have been validated by experimental findings. There is thus ample room for theorists to pick and choose the findings that support their particular point of view, and on that basis favour one or the other model. Many authors seem to have taken the approach that if their findings favour one model, then this disproves the other. In the context of allowing the marketplace of ideas to flourish, this rhetorical (some would say argumentative) approach is not necessarily a bad thing. But at some point a reckoning becomes useful, and that is our aim in the present contribution. Let us

therefore review some of the recent findings that are most pertinent before discussing possible resolutions of these seemingly irreconcilable models.

## 2. EVIDENCE THAT SIMPLE ACOUSTIC FEATURES OF SOUNDS CAN EXPLAIN PATTERNS OF HEMISPHERIC SPECIALIZATION

One of the challenges in studying speech is that it is an intricately complex signal and contains many different acoustical components that carry linguistic, paralinguistic and non-linguistic information. Many early behavioural studies focused on particular features of speech (such as formant transitions and voice-onset time) that distinguish stop consonants from one another, and found that these stimuli were perceived differently from most other sounds because they were parsed into categories. Furthermore, discrimination within categories was much worse than across categories, which violates the more typical continuous perceptual function that is observed with non-speech sounds (Eimas 1963). This categorical perception was believed to be the hallmark of the speech perception 'mode'. A particularly compelling observation, for instance, was made by Whalen & Liberman (1987), who noted that the identical formant transition could be perceived categorically or not depending on whether it formed part of a sound complex perceived as a speech syllable or not. This sort of evidence suggested that the physical cues themselves are insufficient to explain the perceptual categorization phenomena.

However, much other research showed that, at least in some cases, invariant acoustical cues did exist for phonetic categories (Blumstein 1994), obviating the need for a special speech-unique decoding system. Moreover, it was found that categorical perception was not unique to speech because it could be elicited with non-speech stimuli that either emulated certain speech cues (Miller *et al.* 1976; Pisoni 1977), or were based on learning of arbitrary categories, such as musical intervals (Burns & Ward 1978; Zatorre 1983). Eimas *et al.* (1971), together with other investigators (Kuhl 2000), also demonstrated that infants who lacked the capacity to articulate speech could nonetheless discriminate speech sounds in a manner similar to adults, casting doubt on the link between perception and production. The concept that speech may not depend upon a unique human mechanism but rather on general properties of the auditory system was further supported by findings that chinchillas and quail perceived phonemes categorically (Kuhl & Miller 1975; Kluender *et al.* 1987). Thus, from this evidence, speech came to be viewed as one class of sounds with certain particular properties, but not requiring some unique system to handle it.

A parallel trend can be discerned in the neuroimaging literature dealing with the neural basis of speech: particular patterns of neural engagement have been observed in many studies, which could be taken as indicative of specialized speech processors, but other studies have shown that non-speech sounds can also elicit the same patterns. For example, consider studies dealing with the specialization of auditory cortical areas in the left hemisphere. Many of the first functional

neuroimaging studies published were concerned with identifying the pathways associated with processing of speech sounds. Most of these studies did succeed in demonstrating greater response from left auditory cortical regions for speech sounds as compared with non-speech controls, such as tones (Binder *et al.* 2000*a*), amplitude-modulated noise (Zatorre *et al.* 1996) or spectrally rotated speech (Scott *et al.* 2000). However, as noted above, it is somewhat difficult to interpret these responses owing to the complexity of speech; thus, if speech sounds elicit a certain activity pattern that is not observed with some control sound, it is not clear whether to attribute it to the speech *qua* speech, or to some acoustical feature contained within the speech signal but not in the control sound. Conversely, if a non-speech sound that is akin to speech elicits left auditory cortical activation, then one can always argue that it does so by virtue of its similarity to speech. Such findings can therefore comfort adherents of both models.

More recently, however, there has been greater success among functional imaging studies that have focused on the specific hypothesis that rapidly changing spectral energy may require specialized left auditory cortical mechanisms independently of whether they are perceived as speech. One such study was carried out by Belin *et al.* (1998) who used positron emission tomography (PET) to examine processing of formant transitions of different durations in pseudospeech syllables. The principal finding was that whereas left auditory cortical response was similar to both slower and faster transitions, the right auditory cortex responded best to the slower transitions, indicating a differential sensitivity to speed of spectral change. Although the stimuli used in this study were not speech, they were derived from speech signals and therefore one could interpret the result in that light. Such was not the case with a PET experiment by Zatorre & Belin (2001), who used pure-tone sequences that alternated in pitch by one octave at different temporal rates. As the speed of the alternation increased, so did the neural response in auditory cortices in both hemispheres, but the magnitude of this response was significantly greater on the left than on the right. Although this observation is different from the Belin *et al.* (1998) result, which showed little difference in left auditory cortex between faster and slower transitions, it supports the general conclusion of greater sensitivity to temporal rate on the left, using stimuli that bear no relationship to speech sounds. The findings of Zatorre & Belin were recently replicated and extended by Jamison *et al.* (2006), who used functional magnetic resonance imaging (fMRI) with the same stimuli, and found very consistent results even at an individual subject level. A further test of the general proposition is provided in a recent fMRI study by Schönwiesner *et al.* (2005), who used a sophisticated stimulus manipulation consisting of noise bands that systematically varied in their spectral width and temporal rate of change (figure 1). The findings paralleled those of the prior studies to the extent that increasing rate of change elicited a more consistent response from lateral portions of Heschl's gyrus on the left when compared with the right. Although the

precise cortical areas identified were somewhat different from those of the other studies, no doubt related to the very different stimuli used, the overall pattern of lateralization was remarkably similar.

A series of additional studies have also recently been carried out supporting this general trend. For example, Zaehle *et al.* (2004) carried out an fMRI study comparing the activation associated with speech syllables that differed in voice-onset time, and non-speech noises that differed in gap duration. Both classes of stimuli vary in terms of the duration between events, but in one case it cues a speech-relevant distinction, and the stimuli are perceived as speech, whereas the other stimuli are merely broadband noises with a certain size gap inserted. The findings indicated that there was a substantial overlap within left auditory cortices in the response to both speech syllables and noises (figure 2). Thus, the physical cues present in both stimuli, regardless of linguistic status, seemed to be the critical factor determining recruitment of left auditory cortex, as predicted by the cue-specific hypothesis. Further consistent findings were reported by another research group (Joanisse & Gati 2003) who contrasted speech versus non-speech along with slow versus fast changes in a  $2 \times 2$  factorial design. The most relevant finding was that certain areas of the superior temporal gyrus (STG) in both hemispheres responded similarly to speech and non-speech tone sweeps that both contained rapidly changing information; however, the response was substantially greater in the left hemisphere. No such response was observed to stimuli with more slowly changing temporal information, whether speech (vowels) or non-speech (steady-state tones). Thus, in this study too, the physical cues present in the stimuli seem to predict left auditory cortex activation, regardless of whether the stimuli were speech or not.

A final, recent, example of hemispheric asymmetries in auditory cortices arising from low-level features is provided by Boemio *et al.* (2005), who parametrically varied the segment transition rates in a set of concatenated narrowband noise stimuli, such that the segment durations varied from quickly changing (12 ms) to more slowly changing (300 ms). Sensitivity to this parameter was demonstrated bilaterally in primary and adjacent auditory cortices and, unlike the studies just reviewed, the strength of the response was essentially identical on both sides. However, in more downstream auditory regions within the superior temporal sulcus (STS), a clear asymmetry was observed, with the more slowly modulated signals preferentially driving the regions on the right side. The authors conclude that '...there exist two timescales in STG...with the right hemisphere receiving afferents carrying information processed on the long time-scale and the left hemisphere those resulting from processing on the short time-scale.' and that their findings are '...consistent with the proposal suggesting that left auditory cortex specializes in processing stimuli requiring enhanced temporal resolution, whereas right auditory cortex specializes in processing stimuli requiring higher frequency resolution' (Boemio *et al.* 2005, p. 394). Thus, the conclusion reached by an impressive number of independent studies is that these low-level features drive hemispheric differences.



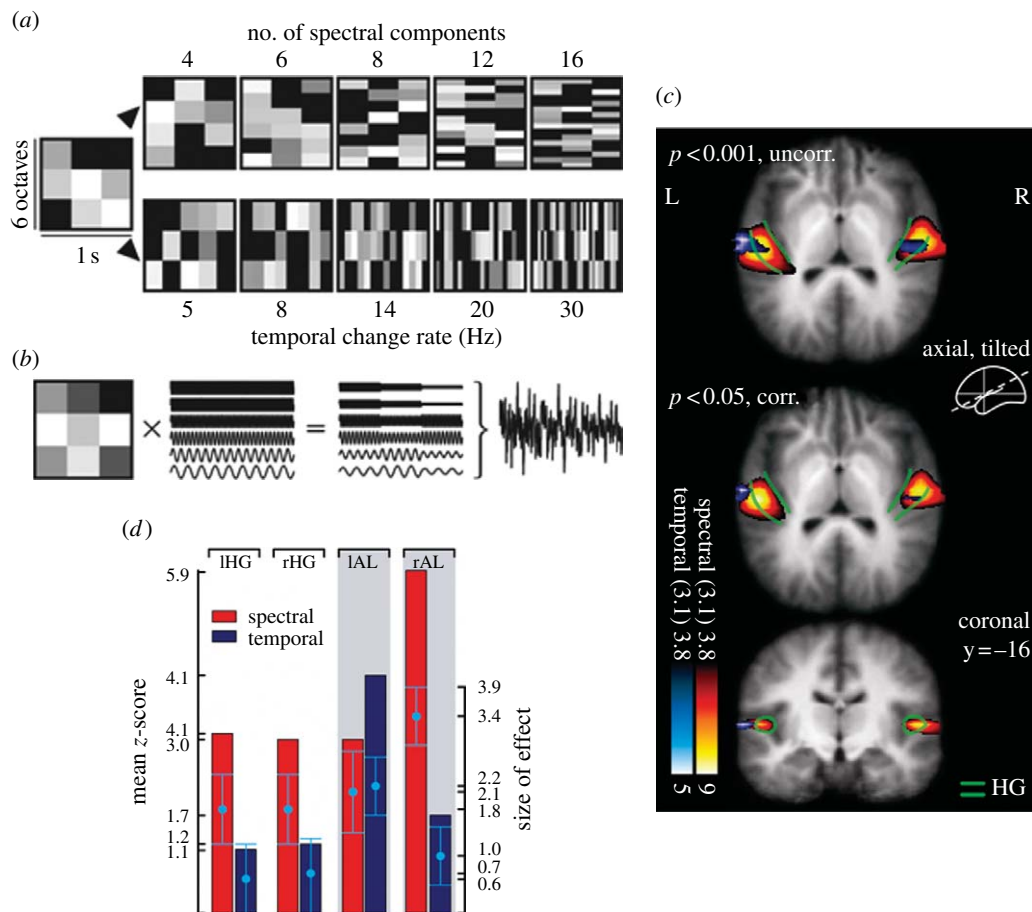


Figure 1. Hemispheric differences in auditory cortex elicited by noise stimuli. (a,b) Illustration of how noise stimuli were constructed; each matrix illustrates stimuli with different bandwidths (on the ordinate) and different temporal modulation rates (on the abscissa). (c) fMRI results indicating bilateral recruitment of distinct cortical areas for increasing rate of temporal or spectral modulation. (d) Effect sizes in selected areas of right (r) and left (l) auditory cortices. Note significant interaction between left anterolateral region, which responds more to temporal than to spectral modulation, and right anterolateral region, which responds more to spectral than to temporal modulation. HG, Heschl's gyrus (Schönwiesner *et al.* 2005).

One might be concerned with the substantial differences across these studies in terms of precisely which cortical zones are recruited, and which ones show hemispheric differences: core or adjacent belt cortices in some studies (Zatorre & Belin 2001; Schönwiesner *et al.* 2005; Jamison *et al.* 2006), as opposed to belt or parabelt areas in STS or anterior STG in others (Joanisse & Gati 2003; Boemio *et al.* 2005). Also, one might point out discrepancies of the specific circumstances under which one observes either enhanced left-hemisphere response to rapidly changing stimuli (Zaehle *et al.* 2004) or, instead, preferential right-hemisphere involvement for slowly varying stimuli (Belin *et al.* 1998; Boemio *et al.* 2005) or, in several cases, both simultaneously (Schönwiesner *et al.* 2005; Jamison *et al.* 2006). Indeed, these details remain to be specified in any clear way, especially in terms of which pathways in the auditory processing stream are being engaged in the various studies (Hickok & Poeppel 2000; Scott & Johnsrude 2003). One might also object, as Scott & Wise (2004) have, that not every study that has examined hemispheric differences based on low-level cues has succeeded in finding them (Giraud *et al.* 2000; Hall *et al.* 2002), or have found them, but not in auditory cortex (Johnsrude *et al.* 1997). There are many reasons that one could

generate why a particular study may not have yielded significant hemispheric differences or observed them in predicted regions, and this too remains to be worked out in detail. But what is striking in the papers reviewed above is the consistency of the overall pattern, as observed repeatedly by a number of different research groups despite the very different stimuli and paradigms used. In particular, we call attention to the clear, replicable evidence that stimuli which are not remotely like speech either perceptually or acoustically (e.g. Boemio *et al.* 2005; Schönwiesner *et al.* 2005; Jamison *et al.* 2006) can and do elicit patterns of hemispheric differences that are consistent and predictable based on acoustical cues, and that brain activation areas are often, though not exclusively, overlapping with those elicited by speech within the left auditory cortex.

Furthermore, the conclusions from these imaging studies fit in a general way with earlier behavioural-lesion studies that focused on the idea that temporal information processing may be especially important for speech. Among the earliest to argue in favour of this idea were researchers working with aphasic populations, who noted associations between aphasia and temporal judgement deficits (Efron 1963; Swisher & Hirsh 1972; Phillips & Farmer 1990; von Steinbüchel

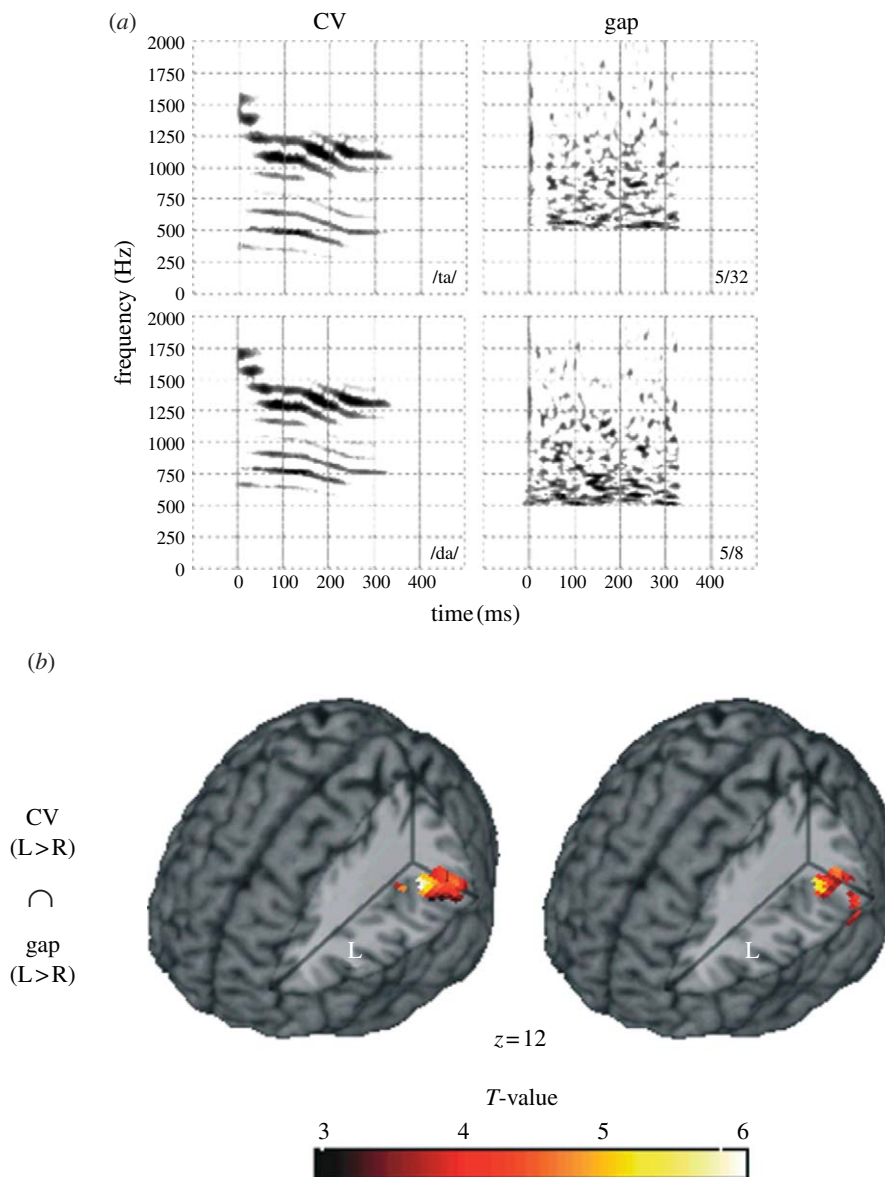


Figure 2. (a) Illustrations of consonant–vowel (CV) speech and non-speech sounds containing similar acoustical properties. (b) fMRI brain images illustrating an overlap between left auditory cortex responses to syllables and to non-speech gap stimuli in left auditory cortex. Similar results are obtained at two distinct time points (TPA1 and TPA2; Zaehe *et al.* 2004).

1998), and from research on children with specific language impairment, who seem to demonstrate global temporal processing deficits (Tallal *et al.* 1993, 1996). A similar conclusion regarding the perceptual deficits of patients with pure word deafness was reached by Phillips & Farmer (1990), who commented that the critical problem in these patients relates to a deficit in processing of sounds with temporal content in the milliseconds to tens of milliseconds range. Complementary data pointing in the same direction are provided by depth electrode recordings from human auditory cortex (Liégeois-Chauvel *et al.* 1999). These authors observed that responses from left but not right Heschl's gyrus distinguished differences related to the voice-onset time feature in speech; critically, a similar sensitivity was present for non-speech analogues that contained similar acoustic features, supporting the contention that it is general temporal acuity which is important, regardless of whether the sound is speech or not. Thus, the neuroimaging findings, which we focus

on here, are, broadly speaking, also compatible with a wider literature.

The studies reviewed above, which argue for the importance of low-level temporal properties of the acoustical signal in defining the role of left auditory cortices, have also led to some theoretical conclusions that have implications beyond the debate about speech. In particular, a major advantage of the cue-specific hypothesis is that because it is neutral about the more abstract status of the stimulus, it can make predictions about all classes of signals, not just speech. Another advantage of the cue-specific model is that it leads to more direct hypotheses about the neural mechanisms that may be involved.

For example, both Zatorre *et al.* (2002) and Poeppel (2003) have independently proposed models whereby differences in neural responses between the left and right auditory cortices are conceptualized as being related to differences in the speed with which dynamically changing spectral information is

processed. The Poeppel model emphasizes differential time integration windows, while Zatorre and colleagues emphasize the idea of a relative trade-off in temporal and spectral resolution between auditory cortices in the two hemispheres. One useful feature of these models is that they generate testable predictions outside the speech domain. The domain of tonal processing is especially relevant in this respect.

### 3. EVIDENCE FOR RIGHT AUDITORY CORTEX PROCESSING OF PITCH INFORMATION

A large amount of evidence has accumulated, indicating that right auditory cortex is specialized for the processing of pitch information under specific circumstances. In neuroimaging studies, asymmetric responses favouring right auditory cortices have been reported in tasks that require pitch judgments within melodies (Zatorre *et al.* 1994) or tones (Binder *et al.* 1997); maintenance of pitch while singing (Perry *et al.* 1999); imagery for familiar melodies (Halpern & Zatorre 1999), or for the timbre of single tones (Halpern *et al.* 2004); discrimination of pitch and duration in short patterns (Griffiths *et al.* 1999); perception of melodies made of iterated ripple noise (Patterson *et al.* 2002); reproduction of tonal temporal patterns (Penhune *et al.* 1998); timbre judgments in dichotic stimuli (Hugdahl *et al.* 1999); perception of missing fundamental tones (Patel & Balaban 2001); and detection of deviant chords (Tervaniemi *et al.* 2000), to cite a few.

What is of interest is that these findings can be accounted for at least in part by the same set of assumptions of a cue-specific model in terms of differential temporal integration. The concept is that encoding of pitch information will become progressively better the longer the sampling window; this idea follows from straightforward sampling considerations—the more cycles of a periodic signal that can be integrated, the better the frequency representation should be. Hence, if the right auditory cortex is proposed to be a slower system—a disadvantage presumably when it comes to speech analysis—it would conversely have an advantage in encoding information in the frequency domain. Several of the studies reviewed in §2 explicitly tested this idea as a means of contrasting the response of the two auditory cortices. For example, Zatorre & Belin (2001) and Jamison *et al.* (2006) demonstrated that neural populations in the lateral aspect of Heschl's gyrus bilaterally responded to increasingly finer pitch intervals in a stimulus, and that this response was greater in the right side. Schönwiesner *et al.* (2005) found a similar phenomenon, but notably their stimuli contained no periodicity, since they used only noise bands of different filter widths. Nonetheless, the right auditory cortical asymmetry emerged as bandwidth decreased, supporting the idea that spectral resolution is greater in those regions.

Additional evidence in favour of the importance of right auditory cortical systems in the analysis and encoding of tonal information comes from a number of other sources, notably behavioural–lesion studies which have consistently reported similar findings.

Damage to superior temporal cortex on the right affects a variety of tonal pitch processing tasks (Milner 1962; Sidtis & Volpe 1988; Robin *et al.* 1990; Zatorre & Samson 1991; Zatorre & Halpern 1993; Warrier & Zatorre 2004). More specifically, lesions to right but not left auditory cortical areas within Heschl's gyrus specifically impair pitch-related computations, including the perception of missing fundamental pitch (Zatorre 1988), and direction of pitch change (Johnsrude *et al.* 2000). The latter study is particularly relevant for our discussion for two reasons. First, the effect was limited to lesions encroaching upon the lateral aspect of Heschl's gyrus—the area seen to respond to small frequency differences by Zatorre & Belin (2001)—and not seen after more anterior temporal lobe damage. Second, the lesion resulted not in an abolition of pitch-discrimination ability, but rather in a large increase in the discrimination threshold. This result fits with the idea of a *relative* hemispheric asymmetry related to resolution. Auditory cortices on both sides must therefore be sensitive to information in the frequency domain, but the right is posited to have a finer resolution; hence, lesions to this region result in an increase in the minimum frequency needed to indicate the direction of change.

Taken together, the neuroimaging findings and the lesion data just reviewed point to a clear difference favouring the right auditory cortex in frequency processing, a phenomenon which can be explained based on a single simplifying assumption that differences exist in the capacity of auditory cortices to deal with certain types of acoustical information. The parsimony of making a single assumption to explain a large body of data concerning processing of frequency information as well as temporal information relevant to speech is attractive. Based on the foregoing empirical data, it seems impossible to support the proposition that low-level acoustical features of sounds have *no* predictive power with respect to patterns of hemispheric differences in auditory cortex. Yet, if one accepts this conclusion, it need not necessarily follow that higher-order abstract features of a stimulus would have no bearing on the way that they are processed in the brain, nor that low-level features can explain all aspects of neural specialization. In fact, despite the clear evidence just mentioned regarding tonal processing, the story becomes much more complex (and interesting) when tonal cues become phonemic in a speech signal, as we shall see in §§7–12. Before reviewing information on use of tonal cues in speech, however, we review some examples of findings that cannot be explained purely on the basis of the acoustic features present in a stimulus, and show that more abstract representations, or context effects, play an important role as well.

### 4. EVIDENCE THAT LINGUISTIC FEATURES OF A STIMULUS CAN INFLUENCE PATTERNS OF HEMISPHERIC SPECIALIZATION

In a fundamental way, a model that does not take into account the linguistic status of a speech sound at all in predicting neural processing pathways cannot be complete; indeed, consider that if a sound which is



speech was processed *identically* to a sound which is not speech at all levels in the brain, then we should not ever be able to perceive a difference! The fact that a speech sound is interpreted as speech, and may lead to phonetic recognition, or to retrieval of meaning, immediately implies contact with memory traces; hence, on those grounds alone different mechanisms would be expected. For instance, studies by [Scott \*et al.\* \(2000\)](#) and [Narain \*et al.\* \(2003\)](#) demonstrate that certain portions of the left anterior and posterior temporal lobes respond to intelligible speech sentences, whether they are produced normally or generated via a noise vocoding algorithm which results in an unusual, unnatural timbre but can still be understood. This result can be understood as revealing the brain areas in which meaning is processed based not upon the acoustical details of the signal, but rather on higher-order processes involved with stored phonetic and/or semantic templates. However, this effect need not indicate that auditory cortices earlier on in the processing stream are not sensitive to low-level features. Rather, it indicates a convergence of processing for different stimulus types at higher levels of analysis where meaning is decoded.

As noted above, adjudicating between the competing models is not always easy because differences in neural activity that may be observed for speech versus non-speech sounds are confounded by possible acoustical differences that are present in the stimuli. One clever way around this problem is to use a stimulus that can be perceived as speech or not under different circumstances. Sine-wave speech provides just such a stimulus, because it is perceived by naive subjects as an unusual meaningless sound, but, after some training, it can usually be perceived to have linguistic content ([Remez \*et al.\* 1981](#)). It has been shown that sine-wave speech presented to untrained subjects does not result in the usual left auditory cortex lateralization seen with speech sounds ([Vouloumanos \*et al.\* 2001](#)), a result taken as evidence for speech specificity. However, sine-wave stimuli are not physically similar to real speech, and thus the differences may still be related to acoustical variables. Two recent fMRI studies exploit sine-wave speech by comparing how these sounds are perceived before and after training sessions, which resulted in a subject being able to hear speech content ([Dehaene-Lambertz \*et al.\* 2005](#); [Möttönen \*et al.\* 2006](#)). In both studies, the principal finding was that cortical regions in the left hemisphere showed enhanced activity after training. In the Dehaene-Lambertz study, however, significant left-sided asymmetries were noted even before training. In the Möttönen study, the enhanced lateralization was seen only in those subjects who were able to learn to identify the stimuli as speech. Thus, these findings demonstrate with an elegant paradigm that identical physical sounds are processed differently when they are perceived as speech than when they are not, although there is also evidence that even without training they may elicit lateralized responses.

A similar but more specific conclusion was reached in a recent study directly examining the neural correlates of categorical perception ([Liebenthal \*et al.\* 2005](#)). These authors compared categorically perceived speech

syllables with stimuli containing the same acoustical cues, but which are perceived neither as speech nor categorically. Their finding of a very clear left STS activation exclusively for the categorically perceived stimuli indicates that this region responds to more than just the acoustical features, since these were shared across the two types of sounds used. The authors conclude that this region performs an intermediate stage of processing linking early processing regions with more anteroventral regions containing stored representations. More generally, these results indicate once again that experience with phonetic categories, and not just physical cues, influences patterns of activity, a conclusion which is also consistent with cross-language studies using various methodologies. For example, [Näätänen and colleagues](#) showed that the size of the mismatch negativity (MMN) response in the left hemisphere is affected by a listener's knowledge of vowel categories in their language ([Näätänen \*et al.\* 1997](#)). [Golestani & Zatorre \(2004\)](#) found that several speech-related zones, including left auditory cortical areas, responded to a greater degree after training with a foreign speech sound than they had before; since the sound had not changed, it is clearly the subjects' knowledge that caused additional activation.

## 5. EVIDENCE THAT CONTEXT EFFECTS MODULATE EARLY SENSORY CORTICAL RESPONSE

There is a related body of empirical evidence, deriving from studies of task-dependent modulation, or context effects, that also indicates that the acoustical features of a stimulus by themselves are insufficient to explain all patterns of hemispheric involvement. Interestingly, such effects have been described for both speech and non-speech stimuli. Thus, they are not predicted by a strong form of either model, since, according to a strict cue-specific model, only stimulus features and not task demands are relevant, while the domain-specific model focuses on the idea that speech is processed by a dedicated pathway, but makes no predictions for non-speech sounds.

For instance, a recent study ([Brechmann & Scheich 2005](#)) using frequency-modulated tones shows that an asymmetry favouring the right auditory cortex emerges for these stimuli only when subjects are actively judging the direction of the tone sweep, and not under passive conditions. In a second condition of this study, these investigators contrasted two tasks, one requiring categorization of the duration of the stimulus, while the other required categorization of the direction of pitch change, using identical stimuli in both conditions. One region of posterior auditory cortex was found to show sensitivity to the task demands, such that more activation was noted in the left than the right for the duration task, while the opposite lateralization emerged for the pitch task. While this result is not predicted by a strict bottom-up model, it is, broadly speaking, consistent with the cue-specific model described above according to which pitch information is better processed in the right auditory cortex while temporal properties are better analysed in the left auditory cortex. However, this model would need to be refined

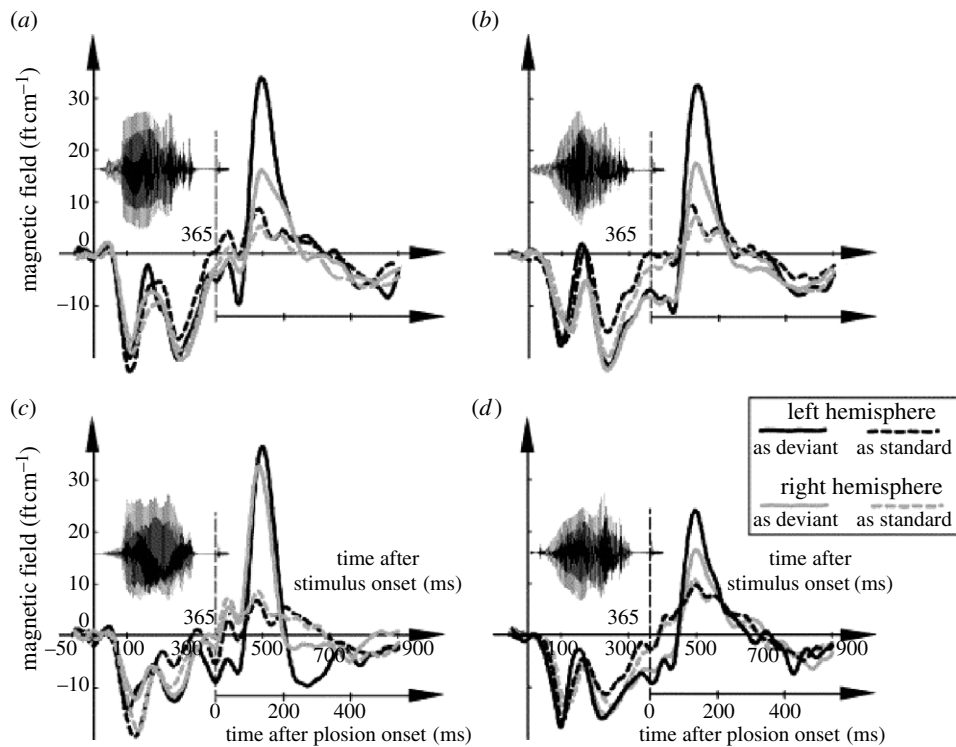


Figure 3. Magnetic-evoked potentials to a non-speech sound presented in different contexts. A left-hemisphere advantage is observed only when the target sound is embedded in (a,b) a real word (verb or noun, respectively), but not when it is presented in either (c) a non-speech context or (d) a pseudoword context, where it is perceived as speech but not as a recognizable word (Shtyrov *et al.* 2005).

in order to take into account not only the nature of the acoustical features present in the stimulus, but also their relevance to a behavioural demand.

Another recent example demonstrating interactions between physical features of a stimulus and context effects comes from a study using magnetoencephalography and a passive-listening MMN paradigm (Shtyrov *et al.* 2005). These investigators used a target sound containing rapidly changing spectral energy, which in isolation is heard as non-speech, in different contexts: one where it is perceived as a speech phoneme and another as a non-speech sound. In addition, the context itself was either a meaningful word or a meaningless but phonetically correct pseudoword. Left-lateralized effects were only observed when the target sound was presented within a word context, and not when it was placed within a pseudoword context, even though it was perceived as speech in the latter case (figure 3). This result is important because, once again, it is not predicted by a strong form of either model. A domain-specific prediction would be that as soon as the sound is perceived as speech, it should result in recruitment of left auditory cortical speech zones, but this was not observed. Similarly, since the target sound is always the same, a strict cue-specific model would have to predict similar patterns regardless of context. Instead, the findings of this study point to the interaction between acoustical features and learned representations.

## 6. EVIDENCE THAT NOT ALL LANGUAGE PROCESSING INVOLVES ACOUSTICAL CUES: SIGN LANGUAGE STUDIES

One of the most dramatic demonstrations of the independence of abstract, symbolic processing involved

in language from the low-level specializations relevant to speech comes from the study of sign language. Since perception of sign languages is purely visual, then whatever results one obtains with them must perforce pertain to general linguistic properties and be unrelated to auditory processing. The existing literature on sign language aphasia in deaf persons suggests that left-hemisphere damage is associated with aphasia-like syndromes, manifested as impairments in signing or in perceiving sign (Hickok *et al.* 1998). More recently, this literature has been enhanced considerably by functional neuroimaging studies of neurologically intact deaf signers. A complete review of this expanding literature is not our aim here, but there are some salient findings which are very relevant to our discussion. In particular, there is a good consensus on the conclusions that (i) what would typically be considered auditory cortex can be recruited for visual sign language processing in the deaf, and (ii) under many circumstances, sign language processing recruits left-hemisphere structures. Moreover, these findings are consistent across a range of distinct sign languages, including American, Quebec, British and Japanese sign languages.

The finding that auditory cortex can be involved in the processing of visual sign can perhaps best be interpreted in light of cross-modal plasticity effects. Many studies have now shown that visual cortex in the blind is functional for auditory (Kujala *et al.* 2000; Weeks *et al.* 2000; Gougoux *et al.* 2005) and tactile tasks (Sadato *et al.* 1996). It is therefore reasonable to conclude that the cross-modal effects in the deaf are not necessarily related to language, a conclusion strengthened by the finding that superior temporal areas respond to non-linguistic visual stimuli in the deaf (Finney *et al.* 2001). But the



recruitment of left-hemisphere language areas for the performance of sign language tasks would appear to be strong evidence in favour of a domain-specific model (Bavelier *et al.* 1998). For example, Petitto *et al.* (2000) showed that when deaf signers are performing a verb generation task, neural activity is seen in the left inferior frontal region, comparable with what is observed with speaking subjects (Petersen *et al.* 1988). Similarly, MacSweeney *et al.* (2002) demonstrated that left-hemisphere language areas in frontal and temporal lobes are recruited in deaf signers during a sentence comprehension task, and that these regions overlap with those of hearing subjects performing a similar speech-based task. However, the latter study also points out that left auditory cortical areas are more active for speech than for sign in hearing persons who sign as a first language; therefore, the left auditory cortex would seem to have some privileged status for processing speech.

In any case, these findings and others would seem to spell trouble for a cue-specific model that predicts hemispheric differences based on acoustical features. The logic goes something like this: if visual signals can result in recruitment of left-hemisphere speech zones, then how can rapid temporal processing of sound be at all relevant? At first glance, this would appear to be strong evidence in favour of a domain-specific model. However, this reasoning does not take into account an evolutionary logic. That is, anyone born deaf still carries in his or her genome the evolutionary history of the species, which has resulted in brain specializations for language. What has changed in the deaf individual is the access to these specializations, since there is a change in the nature of the peripheral input. But it was processing of *sound* that presumably led to the development of the specializations during evolutionary history, assuming that the species has been using speech and not sign during its history (indeed, there is no evidence of any hearing culture anywhere not using speech for language communications). What the literature on deaf sign language processing teaches us is certainly something important that higher-order linguistic processes can be independent of specializations for acoustical processing. But, as with the literature discussed above, this conclusion does not mean that the cue-specific model is incorrect; instead, it tells us, as do many of the other findings we have reviewed, that there are interesting and complex interactions between low-level sensory processing mechanisms and higher-order abstract processing mechanisms.

## 7. EVIDENCE FOR AN INTERACTION BETWEEN PROCESSING OF PITCH INFORMATION AND HIGHER-ORDER LINGUISTIC CATEGORIES: TONAL LANGUAGES

As we have seen, there is consistent evidence for a role of right-hemisphere cortical networks in the processing of tonal information. But tonal information can also be part of a linguistic signal. The question arises whether discrimination of *linguistically relevant* pitch patterns would depend on processing carried out in right-hemisphere networks, as might be predicted by the cue-specific hypothesis, or, alternatively, whether it would engage left-hemisphere speech zones as predicted by

the domain-specific hypothesis. Pitch stimuli that are linguistically relevant at the syllable level thus afford us a unique window for investigating the case for the domain-specific hypothesis.

Tone languages exploit variations in pitch at the syllable level to signal differences in word meaning (Yip 2003). The bulk of information on tonal processing in the brain comes primarily from two tone languages: Mandarin Chinese, which has four contrastive tones, and Thai, which has five (for reviews, see Gandour (1998, 2006a,b)). Similar arguments in support of domain-specificity could be made for those languages in which duration is used to signal phonemic oppositions in vowel length (Gandour *et al.* 2002a,b; Minagawa-Kawai *et al.* 2002; Nenonen *et al.* 2003). However, cross-language studies of temporal processing do not provide a comparable window for adjudicating between cue- and domain-specificity, because both temporal and language processes are generally considered to be mediated primarily by neural mechanisms in the left hemisphere.

## 8. EVIDENCE FOR LEFT-HEMISPHERE INVOLVEMENT IN PROCESSING OF LINGUISTIC TONE

As reviewed in §§2 and 3, in several functional neuroimaging studies of phonetic and pitch discrimination in speech, right prefrontal cortex was activated in response to pitch judgments for English-speaking listeners, whereas consonant judgments elicited activation of the left prefrontal cortex (Zatorre *et al.* 1992, 1996). In English, consonants are phonemically relevant; pitch patterns at the syllable level are not. What happens when pitch patterns are linguistically relevant in the sense of cuing lexical differences? Several studies have now demonstrated a left-hemisphere specialization for tonal processing in native speakers of Mandarin Chinese. For example, when asked to discriminate Mandarin tones and homologous low-pass filtered pitch patterns (Hsieh *et al.* 2001), native Chinese-speaking listeners extracted tonal information associated with the Mandarin tones via left inferior frontal regions in both speech and non-speech stimuli. English-speaking listeners, on the other hand, exhibited activation in homologous areas of the right hemisphere. Pitch processing is therefore lateralized to the left hemisphere only when the pitch patterns are phonologically significant to the listener; otherwise, to the right hemisphere. It appears that left-hemisphere mechanisms mediate processing of linguistic information irrespective of acoustic cues or type of phonological unit, segmental or suprasegmental. No activation occurred in left frontal regions for English listeners on either the tone or pitch task because they were judging pitch patterns that are not phonetically relevant in the English language.

Given these observations, one might then ask whether knowledge of a tone language changes the hemispheric dynamics for processing of tonal information in general, or whether it is specific to one's own language. The answer appears to be that it is tied closely to knowledge of a specific language. When asked to discriminate Thai tones, Chinese listeners fail

to show activation of left-hemisphere regions (Gandour *et al.* 2000, 2002a). This finding is especially interesting inasmuch as the tonal inventory of Thai is similar to that of Mandarin in terms of number of tones and type of tonal contours. In spite of similarities between the tone spaces of Mandarin and Thai, the fact that we still observe cross-language differences in hemispheric specialization of tonal processing suggests that the influence of linguistically relevant parameters of the auditory signal is specific to experience with a particular language. Other cross-language neuroimaging studies of Mandarin have similarly revealed activation of the left posterior prefrontal and insular cortex during both perception and production tasks (Klein *et al.* 1999, 2001; Gandour *et al.* 2003; Wong *et al.* 2004).

An optimal window for exploring how the human brain processes linguistically relevant temporal and spectral information is one in which both vowel length and tone are contrastive. Thai provides a good test case, inasmuch as it has five contrastive tones in addition to a vowel length contrast. To address the issue of temporal versus spectral processing directly, it is imperative that the same group of subjects be asked to make perceptual judgments of tone and vowel length on the same stimuli. To test this idea in a cross-language study (Gandour *et al.* 2002a), Thai and Chinese subjects were asked to discriminate pitch and timing patterns presented in the same auditory stimuli under speech and non-speech conditions. If acoustic in nature, effects due to this level of processing should be maintained across listeners regardless of language experience. If, on the other hand, effects are driven by higher-order abstract features, we expect brain activity of Thai and Chinese listeners to vary depending on language-specific functions of temporal and spectral cues in their respective languages. The question is whether language experience with the same type of phonetic unit is sufficient to lead to similar brain activation patterns as those of native listeners (cf. Gandour *et al.* 2000). In a comparison of pitch and duration judgments of speech relative to non-speech, the left inferior prefrontal cortex was activated for the Thai group only (figure 4). No matter whether the phonetic contrast is signalled primarily by spectral or temporal cues, the left hemisphere appears to be dominant in processing contrastive phonetic features in a listener's native language. Both groups, on the other hand, exhibited similar fronto-parietal activation patterns for spectral and temporal cues under the non-speech condition. When the stimuli are no longer perceived as speech, the language-specific effects disappear. Regardless of the neural mechanisms underlying lower-level processing of spectral and temporal cues, hemispheric specialization is clearly sensitive to higher-order information about the linguistic status of the auditory signal.

The activation of posterior prefrontal cortex in the above-mentioned studies of tonal processing raises questions about its functional role. Because most of these studies used discrimination tasks, there were considerable demands on attention and memory. The influence of higher-order abstract features notwithstanding, the question still remains whether activation

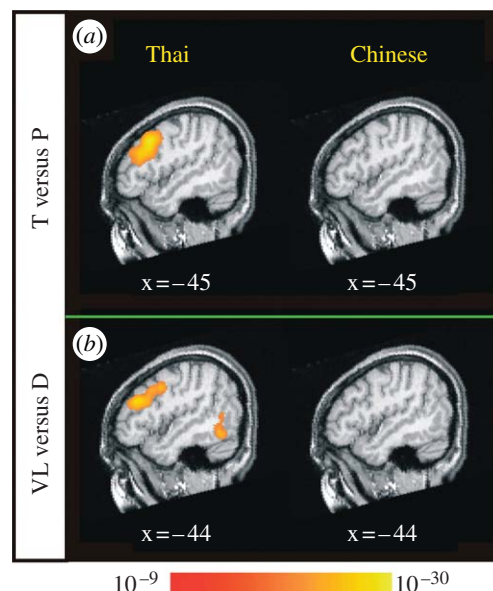


Figure 4. Cortical activation maps comparing discrimination of pitch and duration patterns in a speech relative to a non-speech condition for groups of Thai and Chinese listeners. (a) Spectral contrast between Thai tones (speech) and pitch (non-speech) and (b) temporal contrast between Thai vowel length (speech) and duration (non-speech). Left-sided activation foci in frontal and temporo-occipital regions occur in the Thai group only. T, tone; P, pitch; VL, vowel length; D, duration (Gandour *et al.* 2002a,b).

of this subregion reflects tonal processing, working memory in the auditory modality or other mediational, task-specific processes that transcend the cognitive domain. In an attempt to fractionate mediational components that may be involved in phonetic discrimination, tonal matching was compared with a control condition in which whole syllables were matched to one another (Li *et al.* 2003). The only difference between conditions is the focus of attention, either to a subpart of the syllable or to the whole syllable itself. Selective attention to Mandarin tones elicited activation of a left-sided dorsal fronto-parietal, attention network, including a dorsolateral subregion of posterior prefrontal cortex. This cortical network in the left hemisphere is likely to reflect the engagement of attention-modulated, executive functions that are differentially sensitive to internal dimensions of a whole stimulus regardless of sensory modality or cognitive domain (Corbetta *et al.* 2000; Shaywitz *et al.* 2001; Corbetta & Shulman 2002).

## 9. EVIDENCE THAT TONAL CATEGORIES ARE SUFFICIENT TO ACCOUNT FOR LEFT-HEMISPHERE LATERALITY EFFECTS

Because tones necessarily co-occur with real words in natural speech, it has been argued that cross-language differences in hemispheric laterality reflect nothing more than a lexical effect (Wong 2002). To isolate prelexical processing of tones, a recent study used a novel design in which hybrid stimuli were created by superimposing Thai tones onto Mandarin syllables (*tonal chimeras*) and Mandarin tones onto the same syllables (real words; Xu *et al.* 2006). Chinese and Thai speakers discriminated paired tonal contours in

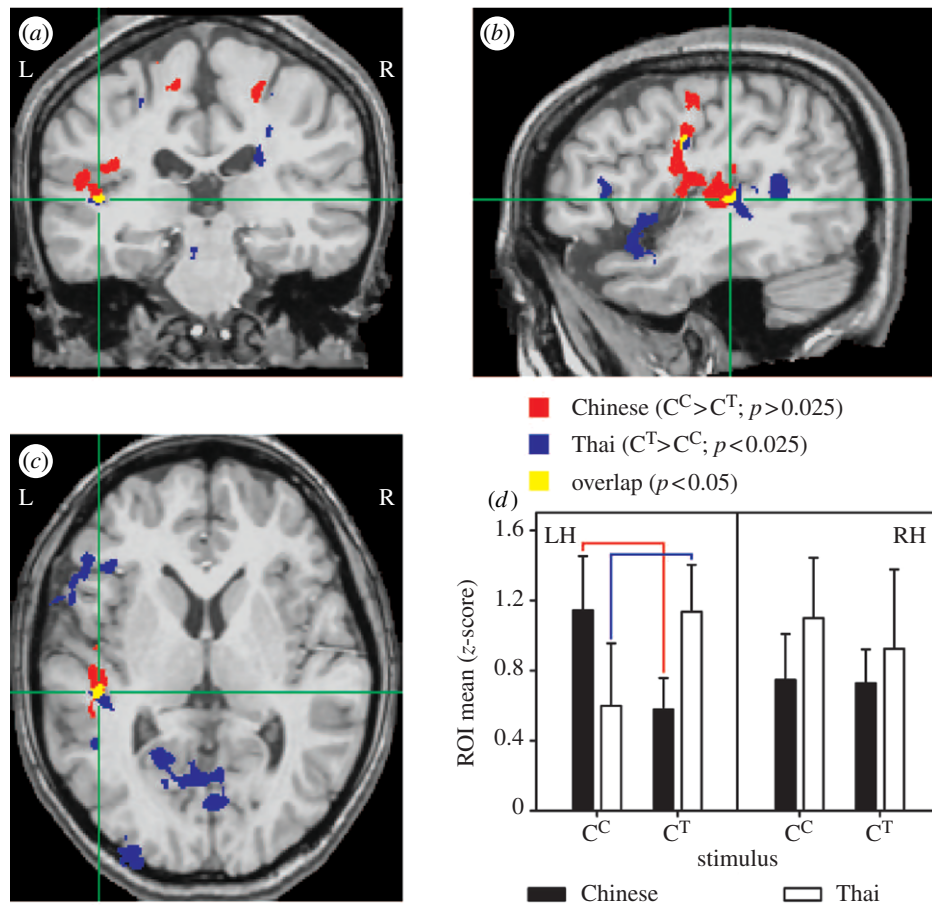


Figure 5. Cortical areas activated in response to discrimination of Chinese and Thai tones. A common focus of activation is indicated by the overlap (yellow) between Chinese and Thai groups in the functional activation maps. Green cross-hair lines mark the stereotactic centre coordinates for the overlapping region in the left planum temporale ((a) coronal section,  $y = -25$ ; (b) sagittal section,  $x = -44$ ; (c) axial section,  $z = +7$ ). A double dissociation ((d) bar charts) between tonal processing and language experience reveals that for the Thai group, Thai tones elicit stronger activity relative to Chinese tones, whereas for the Chinese group, stronger activity is elicited by Chinese tones relative to Thai tones. C<sup>C</sup>, Chinese tones superimposed on Chinese, i.e. Chinese words; C<sup>T</sup>, Thai tones superimposed on Chinese syllable, i.e. *tonal chimeras*; L/LH, left hemisphere; R/RH, right hemisphere; ROI, region of interest (Xu *et al.* 2006).

chimeric and Mandarin stimuli. None of the chimeric stimuli was identifiable as a Thai or Mandarin word. Thus, it was possible to compare Thai listeners' judgments of native tones in tonal chimeras with non-native tones in Mandarin words in the absence of lexical-semantic processing. In a comparison of native versus non-native tones, overlapping activity was identified in the left planum temporale. In this area, a double dissociation between language experience and neural representation of pitch occurred, such that stronger activity was elicited in response to native when compared with non-native tones (figure 5). This finding suggests that cortical processing of pitch information can be shaped by language experience, and, moreover, that lateralized activation in left auditory cortex can be driven by higher-order abstract knowledge acquired through language experience. Both top-down and bottom-up processing are essential features of tonal processing. This reciprocity allows for modification of neural mechanisms involved in pitch processing based on language experience. It now appears that computations at a relatively early stage of acoustic-phonetic processing in auditory cortex can be modulated by the linguistic status of stimuli (Griffiths & Warren 2002).

## 10. EVIDENCE FOR BILATERAL INVOLVEMENT IN PROCESSING SPEECH PROSODY REGARDLESS OF THE LEVEL OF LINGUISTIC REPRESENTATION

In tone languages, pitch contours can be used to signal intonation as well as word meaning. A cross-language (Chinese and English) fMRI study (Gandour *et al.* 2004) was conducted to examine brain activity elicited by selective attention to Mandarin tone and intonation (statement versus question), as presented in three- and one-syllable utterance pairs. The Chinese group exhibited greater activity than the English in the left ventral aspects of the inferior parietal lobule across prosodic units and utterance lengths. It is possible that the 'categoricalness' or phonological relevance of the auditory stimuli triggers activation in this area (Jacquemot *et al.* 2003). In addition, only the Chinese group exhibited a leftward asymmetry in the intraparietal sulcus, anterior/posterior regions of the STG and frontopolar regions. However, both language groups showed activity within the STS and middle frontal gyrus (MFG) of the right hemisphere (figure 6). The rightward asymmetry may reflect shared mechanisms underlying early attentional modulation in processing of complex pitch patterns irrespective of language



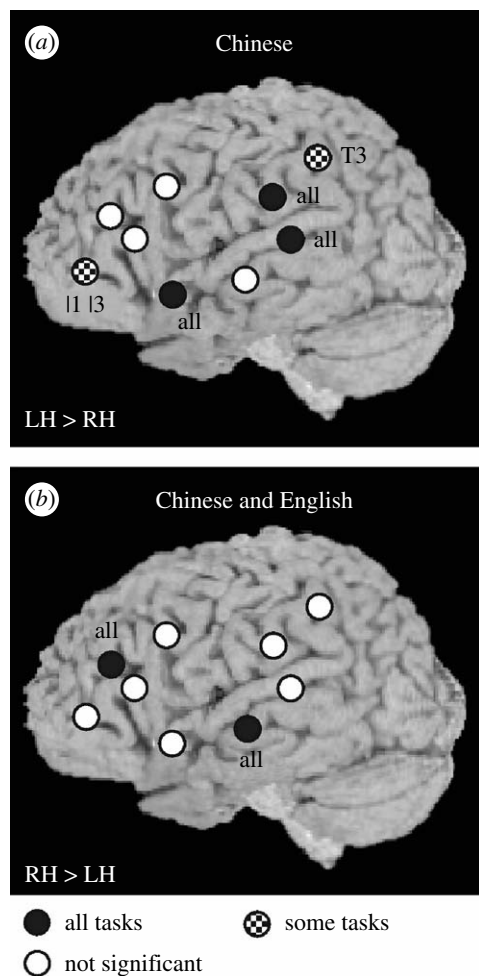


Figure 6. Laterality effects for ROIs in (a) the Chinese group only and in (b) both Chinese and English groups as rendered on a three-dimensional LH template for common reference. (a) In the Chinese group, the ventral aspects of the inferior parietal lobule and anterior and posterior STG are lateralized to the LH across tone and intonation tasks; the anterior MFG and intraparietal sulcus for a subset of tasks. (b) In both groups, the middle portions of the MFG and STS are lateralized to the RH across tasks. This right-sided fronto-temporal network subserves pitch processing regardless of its linguistic function. Other ROIs do not show laterality effects. MFG, middle frontal gyrus; STG, superior temporal gyrus (Gandour *et al.* 2004).

experience (Zatorre *et al.* 1999). Albeit in the speech domain, this fronto-temporal network in the right hemisphere serves to maintain pitch information regardless of its linguistic relevance. Tone and intonation are therefore best thought of as being subserved by a mosaic of multiple local asymmetries, which allows for the possibility that different regions may be differentially weighted in laterality depending on language-, modality- and task-related features. We argue that left-hemisphere activity reflects higher-level processing of internal representations of Chinese tone and intonation, whereas right-hemisphere activity reflects lower-level, domain-independent pitch processing. Speech prosody perception is mediated primarily by the right hemisphere, but lateralized to task-dependent regions in the left hemisphere when language processing is required beyond the auditory analysis of the complex sound.

A cross-language study (Chinese and English) was also conducted to investigate the neural substrates underlying the discrimination of two *sentence-level* prosodic phenomena in Mandarin Chinese: contrastive (or emphatic) stress and intonation (Tong *et al.* 2005). Between-group comparisons revealed that the Chinese group exhibited significantly greater activity in the *left* supramarginal gyrus (SMG) and posterior middle temporal gyrus (MTG) relative to the English group for both tasks. The leftward asymmetry in the SMG and MTG, respectively, is consistent with the notion of a dorsal processing stream emanating from auditory cortex, projecting to the inferior parietal lobule, and ultimately to frontal lobe regions (Hickok & Poeppel 2004) and a ventral processing stream that projects to posterior inferior temporal lobe portions of the MTG and parts of the inferior temporal and fusiform gyri (Binder *et al.* 2000). The involvement of the posterior MTG in sentence comprehension is supported by voxel-based lesion-symptom mapping analysis (Dronkers & Ogar 2004). The left SMG serves as part of an auditory-motor integration circuit in speech perception, and supports phonological encoding-recoding processes in a variety of tasks. For both language groups, *rightward* asymmetries were observed in the middle portion of the MFG across tasks. The rightward asymmetry across tasks and language groups implicates more general auditory attention and working memory processes associated with pitch perception. Its activation is lateralized to the right hemisphere regardless of prosodic unit or language group. This area is not domain-specific since it is similarly recruited for extraction and maintenance of pitch information in processing music (Zatorre *et al.* 1994; Koelsch *et al.* 2002).

These findings from tone languages converge with imaging data on sentence intonation perception in non-tone languages (German and English). In degraded speech (prosodic information only), right-hemisphere regions are engaged predominantly in mediating slow modulation of pitch contours, whereas, in normal speech, lexical and syntactic processing elicits activity in left-hemisphere areas (Meyer *et al.* 2002, 2003). In high memory load tasks that result in recruitment of frontal lobe regions, a rightward asymmetry is found for prosodic stimuli and a leftward asymmetry for sentence processing (Plante *et al.* 2002).

## 11. EVIDENCE FROM SECOND-LANGUAGE LEARNERS FOR BILATERAL INVOLVEMENT IN TONAL PROCESSING

Another avenue for investigating cue- versus domain-specificity is the cortical processing of pitch in a tone language by second-language learners without any previous experience with lexical tones. As pointed out in this review, there is a lack of left-hemisphere dominance in the processing of lexical tone by English speakers who have had no prior experience with a tone language. In the case of lexical tone, there is nothing comparable in English phonology. Thus, a second-language learner of a tone language must develop novel processes for tone perception (Wang *et al.* 2003). The question then arises as to how learning a tone language as a second language affects cortical processing of

pitch. The learning of Mandarin tones by adult native speakers of English was investigated by comparing cortical activation during a tone identification task before and after a two-week training procedure (Wang *et al.* 2003). Cortical effects of learning Chinese tones were associated with increased activation in the left posterior STG and adjacent regions, and recruitment of additional cortical areas within the right inferior frontal gyrus (IFG). These early cortical effects of learning lexical tones are interpreted to involve both the recruitment of pre-existing language-related areas (left STG), consistent with the findings of Golestani & Zatorre (2004) for learning of a non-pitch contrast, as well as the recruitment of additional cortical regions specialized for identification of lexical tone (right IFG). Whereas activation of the left STG is consistent with domain-specificity, the right IFG activation is consistent with cue-specificity. Native speakers of English appear to acquire a new language function, i.e. tonal identification, by enhancing existing pitch processing in the right IFG, a region that heretofore was not specialized for processing linguistic pitch on monosyllabic words. These findings suggest that cortical representations might be continuously modified as learners gain more experience with tone languages.

## 12. EVIDENCE THAT KNOWLEDGE OF TONAL CATEGORIES MAY INFLUENCE EARLY STAGES OF PROCESSING

While it is important to identify language-dependent processing systems at the cortical level, a complete understanding of the neural organization of language can only be achieved by viewing language processes as a set of computations or mappings between representations at different stages of processing (Hickok & Poeppel 2004). From the aforementioned haemodynamic imaging studies at the level of the cerebral cortex, it seems impossible to dismiss early processing stages as not relevant to language processing. Early stages of processing on the input side may perform computations on the acoustic data that are relevant to linguistic as well as non-linguistic auditory perception. However, to best characterize how pitch processing evolves in the time dimension, we need to turn our attention to auditory electrophysiological studies of tonal processing at the cortical level and even earlier stages of processing at the level of the brainstem.

In a study of the role of tone and segmental information in Cantonese word processing (Schirmer *et al.* 2005), the time course and amplitude of the N400 effect, a negativity that is associated with processing the semantic meaning of a word, were comparable for tone and segmental violations (cf. Brown-Schmidt & Canseco-Gonzalez 2004). The MMN, a cortical potential elicited by an odd-ball paradigm, reflects preattentive processing of auditory stimuli. Upon presentation of words with a similar tonal contour (falling) from native (Thai) and non-native (Mandarin) tone languages to Thai listeners (Sittiprapaporn *et al.* 2003), the MMN elicited by a native Thai word (/k<sup>h</sup>a/) was greater than that elicited by a non-native Mandarin word (/ta/) and lateralized to the left auditory cortex. Both the Sittiprapaporn *et al.* (2003) and Schirmer

*et al.* (2005) studies involve spoken word recognition instead of tonal processing *per se*. In a cross-language (Mandarin, English and Spanish) MEG study of spoken word recognition (Valaki *et al.* 2004), the Chinese group revealed a greater degree of late activity (greater than or equal to 200 ms) relative to the non-tone language groups in the right superior temporal and temporo-parietal regions. Since both phonological and semantic processes are involved in word recognition, we can only speculate that this increased RH activity reflects neural processes associated with the analysis of lexical tones.

The degree of linguistic specificity has yet to be determined for computations performed at the level of the auditory brainstem. The conventional wisdom appears to be that ‘although there is considerable ‘tuning’ in the auditory system to the acoustic properties of speech, the processing operations conducted in the relay nuclei of the brainstem and thalamus are general to all sounds, and speech-specific operations probably do not begin until the signal reaches the cerebral cortex’ (Scott & Johnsrude 2003, p. 100). In regard to tonal processing, the auditory brainstem provides a window for examining the effects of the linguistic status of pitch patterns at a stage of processing that is free of attention and memory demands.

To test whether early, preattentive stages of pitch processing at the brainstem level may be influenced by language experience, a recent study (Krishnan *et al.* 2005) investigated the human frequency following response (FFR), which reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem. If based strictly on acoustic properties of the stimulus, FFRs in response to time-varying  $f_0$  contours at the level of the brainstem would be expected to be homogeneous across listeners regardless of language experience. If, on the other hand, FFRs are sensitive to long-term language learning, they may be somewhat heterogeneous depending on how  $f_0$  cues are used to signal pitch contrasts in the listener’s native language. FFRs elicited by the four Mandarin tones were recorded from native speakers of Mandarin Chinese and English. Pitch strength and accuracy of pitch tracking were extracted from the FFRs using autocorrelation algorithms. In the autocorrelation functions, a peak at the fundamental period  $1/f_0$  is observed for both groups, which means that phase-locked activity to the fundamental period is present regardless of language experience. However, the peak for the English group is smaller and broader relative to the Chinese group, suggesting that phase-locked activity is not as robust for English listeners. Autocorrelograms reveal a narrower band phase-locked interval for the Chinese group compared with the English, suggesting that phase-locked activity for the Chinese listeners is not only more robust, but also more accurate in tracking the  $f_0$  contour (figure 7). Both FFR pitch strength and pitch tracking were significantly greater for the Chinese group than for the English across all four Mandarin tones. These data suggest that language experience can induce neural plasticity at the brainstem level that may be enhancing or priming linguistically relevant features of the speech input. Moreover, the prominent phase-locked interval bands

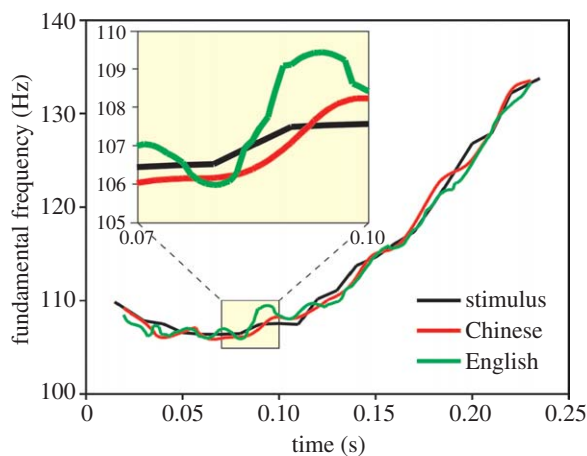


Figure 7. Grand-average  $f_0$  contours of Mandarin tone 2 derived from the FFR waveforms of all subjects across both ears in the Chinese and English groups. The  $f_0$  contour of the original speech stimulus is displayed in black. The enlarged inset shows that the  $f_0$  contour derived from the FFR waveforms of the Chinese group more closely approximates that of the original stimulus (*yi*<sup>2</sup> ‘aunt’) when compared with the English group. (Krishnan *et al.* 2005).

at  $1/f_0$  were similar for stimuli that were spectrally different (speech versus non-speech) but produced equivalent pitch percepts in response to the Mandarin low falling–rising  $f_0$  contour (Krishnan *et al.* 2004). We infer that experience-driven adaptive neural mechanisms are involved even at the level of the brainstem. They sharpen response properties of neurons tuned for processing pitch contours that are of special relevance to a particular language. Language-dependent operations may thus begin before the signal reaches the cerebral cortex.

### 13. CONCLUSION

What have we learned from reviewing this very dynamic and evolving field of research? It should be clear that a conceptualization of hemispheric specialization as being driven by just one simple dichotomy is inadequate. The data we have reviewed show that neural specializations are indeed predictable based on low-level features of the stimulus, and that they can be influenced by linguistic status. The body of evidence cannot be accounted for by claiming one or the other as the sole explanatory model. But how are they to be integrated? A complete account of language processing must allow for multiple dichotomies or scalar dimensions that either apply at different time intervals or interact within the same time interval at different cortical or subcortical levels of the brain. It is not our purpose in this discussion to describe a new model that will solve this problem, but rather to motivate researchers away from a ‘zero-sum game’ mentality.

Each of the models has its strengths and weaknesses. As we have seen, a strong form of either one is untenable. Indeed, several results cannot be predicted based on strong forms of either model. The cue-specific model not only predicts a large body of empirical evidence for both speech and other types of signals, but does so in a parsimonious way; it is best at explaining what is happening in the early stages of

hemispheric processing, and is able to account for both speech and pitch processing. Its principal weakness is that it makes few predictions about later stages and hence does not directly take into account abstract knowledge. What needs to be done to improve it, therefore, is to add to it to take into account the influence of later stages. The domain-specific model as often presented, that is, as a monolithic view that speech is processed by a special mechanism unrelated to any other auditory process, is not very viable. But some account must be given of the many diverse phenomena we have discussed that indicate that higher-order knowledge does indeed influence patterns of neural processing. A less intransigent version of the domain-specific model would simply point out that abstract knowledge is relevant in explaining patterns of neural processing. Such a version can peacefully coexist with a modified version of the cue-specific hypothesis that focuses on early levels of processing and makes no claims about interaction occurring at later stages.

In the case of tonal processing, for example, it appears that a more complete account will emerge from consideration of general sensory-motor and cognitive processes in addition to those associated with linguistic knowledge. For example, the activation of frontal and temporal areas in the right hemisphere reflects their roles in mediating pitch processing irrespective of domain. We have also seen how various executive functions related to attention and working memory mediate prosody perception, in addition to stimulus characteristics, task demands and listeners’ experience. All of these factors will need to be taken into account. Moreover, brain imaging should draw our attention to networks instead of isolated regions; in this respect, we look forward to (i) the further development of the idea that spatio-temporal patterns of neural interactions are key (McIntosh 2000) and (ii) the application of more sophisticated integrative computational models to imaging data (Horwitz *et al.* 2000).

There are several general principles that perhaps should be kept in mind in attempting to forge a solution to the dialectical issue before us. One such principle is that whenever one considers higher-order or abstract knowledge, one necessarily invokes memory mechanisms. There has been insufficient attention to the fact that any domain-specific effect is also a memory effect. As mentioned above, as soon as a sensory signal makes contact with a stored representation, there is going to be an interesting interaction based on the nature of that representation. Neurobiology would lead us to think that abstract memory representations are unlikely to be present in primary sensory cortices. Instead, abstract knowledge is probably present at multiple higher levels of cortical information processing. It is at these levels that many of the language-specific effects are likely to occur. In this respect, it is interesting to note that the effects of such learning can interact all the way down the processing stream, to include subcortical processes, as we have seen.

These kinds of interactive effects bring up a second major principle that is important in elucidating a more comprehensive model of hemispheric processing of speech, and that is the important role of top-down modulation. It is unlikely that early sensory cortices



contain stored representations, but it is probable that the latter can influence early processing via top-down effects. Linguistic status, context effects, learning or attention, can modulate early processing. One promising avenue of investigation therefore would be to elucidate the efferent influences that come from amodal temporal, parietal or frontal cortices and interact with the neural computations occurring in earlier regions. A research effort devoted to understanding this specific sort of interaction is likely to prove fruitful.

In closing, we believe that the past 25 years have produced a very valuable body of evidence bearing on these issues, perhaps motivated by the creative tension inherent in the two contrasting models we have highlighted. If that is indeed the case, then these ideas have served their purpose. But we also believe that now is an appropriate time to move beyond the dichotomy. We look forward to the next 25 years.

We gratefully acknowledge research support from the Canadian Institutes of Health Research (R.J.Z.) and the US Institutes of Health (J.T.G.). We thank our former and present students and collaborators for their contributions to the work and ideas presented herein.

## REFERENCES

- Bavelier, D., Corina, D. P. & Neville, H. J. 1998 Brain and language: a perspective from sign language. *Neuron* **21**, 275–278. (doi:10.1016/S0896-6273(00)80536-X)
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C. & Samson, Y. 1998 Lateralization of speech and auditory temporal processing. *J. Cogn. Neurosci.* **10**, 536–540. (doi:10.1162/089892998562834)
- Binder, J., 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., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J. & Possing, J. 2000 Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528. (doi:10.1093/cercor/10.5.512)
- Blumstein, S. E. 1994 The neurobiology of the sound structure of language. In *The cognitive neurosciences* (ed. M. S. Gazzaniga), pp. 915–929. Cambridge, MA: MIT Press.
- Boemio, A., Fromm, S., Braun, A. & Poeppel, D. 2005 Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* **8**, 389–395. (doi:10.1038/nm1409)
- Brechmann, A. & Scheich, H. 2005 Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb. Cortex* **15**, 578–587. (doi:10.1093/cercor/bhh159)
- Brown-Schmidt, S. & Canseco-Gonzalez, E. 2004 Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *J. Psycholinguist. Res.* **33**, 103–135. (doi:10.1023/B:JOPR.0000017223.98667.10)
- Burns, E. M. & Ward, W. D. 1978 Categorical perception: phenomenon or epiphenomenon. Evidence from experiments in perception of melodic musical intervals. *J. Acoust. Soc. Am.* **63**, 456–468. (doi:10.1121/1.381737)
- Corbetta, M. & Shulman, G. L. 2002 Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215. (doi:10.1038/nrn755)
- Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P. & Shulman, G. L. 2000 Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* **3**, 292–297. (doi:10.1038/73009)
- 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)
- Diehl, R. L., Lotto, A. J. & Holt, L. L. 2004 Speech perception. *Annu. Rev. Psychol.* **55**, 149–179. (doi:10.1146/annurev.psych.55.090902.142028)
- Dronkers, N. & Ogar, J. 2004 Brain areas involved in speech production. *Brain* **127**, 1461–1462. (doi:10.1093/brain/awh233)
- Efron, R. 1963 Temporal perception, aphasia and deja vu. *Brain* **86**, 403–423. (doi:10.1093/brain/86.3.403)
- Eimas, P. D. 1963 The relation between identification and discrimination along speech and nonspeech continua. *Lang. Speech* **6**, 206–217.
- Eimas, P., Siqueland, E., Jusczyk, P. & Vigorito, J. 1971 Speech perception in infants. *Science* **171**, 303–306. (doi:10.1126/science.171.3968.303)
- Finney, E. M., Fine, I. & Dobkins, K. R. 2001 Visual stimuli activate auditory cortex in the deaf. *Nat. Neurosci.* **4**, 1171–1173. (doi:10.1038/nn763)
- Gandour, J. 1998 Aphasia in tone languages. In *Aphasia in atypical populations* (eds P. Coppens, A. Basso & Y. Lebrun), pp. 117–141. Hillsdale, NJ: Lawrence Erlbaum.
- Gandour, J. 2006a Tone: neurophonetics. In *Encyclopedia of language and linguistics* (ed. K. Brown), pp. 751–760, 2nd edn. Oxford, UK: Elsevier.
- Gandour, J. 2006b Brain mapping of Chinese speech prosody. In *Handbook of East Asian psycholinguistics: Chinese* (eds P. Li, L. H. Tan, E. Bates & O. Tzeng), pp. 308–319. New York, NY: Cambridge University Press.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D. & Hutchins, G. D. 2000 A PET cross-linguistic study of tone perception. *J. Cogn. Neurosci.* **12**, 207–222. (doi:10.1162/089892900561841)
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Sathamnuwong, N., Tong, Y. & Li, X. 2002a A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J. Cogn. Neurosci.* **14**, 1076–1087. (doi:10.1162/089892902320474526)
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Sathamnuwong, N., Tong, Y. & Lurito, J. 2002b Neural circuitry underlying perception of duration depends on language experience. *Brain Lang.* **83**, 268–290. (doi:10.1016/S0093-934X(02)00033-0)
- Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Sathamnuwong, N. & Lurito, J. 2003 Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang.* **84**, 318–336. (doi:10.1016/S0093-934X(02)00505-9)
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X. & Lowe, M. 2004 Hemispheric roles in the perception of speech prosody. *Neuroimage* **23**, 344–357. (doi:10.1016/j.neuroimage.2004.06.004)
- Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R. & Kleinschmidt, A. 2000 Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* **84**, 1588–1598.
- Golestani, N. & Zatorre, R. J. 2004 Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* **21**, 494–506. (doi:10.1016/j.neuroimage.2003.09.071)
- Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P. & Lepore, F. 2005 Functional neuroimaging of sound localization with early blind persons. *PLoS Biol.* **3**, 324–333. (doi:10.1371/journal.pbio.0030027)

- Griffiths, T. D. & Warren, J. D. 2002 The planum temporale as a computational hub. *Trends Neurosci.* **25**, 348–353. (doi:10.1016/S0166-2236(02)02191-4)
- Griffiths, T. D., Johnsruide, I. S., Dean, J. L. & Green, G. G. R. 1999 A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport* **10**, 3825–3830. (doi:10.1097/00001756-199912160-00019)
- Hall, D. A., Johnsruide, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A. & Summerfield, A. Q. 2002 Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **12**, 140–149. (doi:10.1093/cercor/12.2.140)
- Halpern, A. R. & 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. (doi:10.1093/cercor/9.7.697)
- Halpern, A. R., Zatorre, R. J., Bouffard, M. & Johnson, J. A. 2004 Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* **42**, 1281–1292. (doi:10.1016/j.neuropsychologia.2003.12.017)
- Hickok, G. & Poeppel, D. 2000 Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131–138. (doi:10.1016/S1364-6613(00)01463-7)
- Hickok, G. & Poeppel, D. 2004 Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99. (doi:10.1016/j.cognition.2003.10.011)
- Hickok, G., Bellugi, U. & Klima, E. 1998 The neural organization of language: evidence from sign language aphasia. *Trends Cogn. Sci.* **2**, 129–136. (doi:10.1016/S1364-6613(98)01154-1)
- Horwitz, B., Friston, K. & Taylor, J. 2000 Neural modeling and functional brain imaging: an overview. *Neural Netw.* **13**, 829–846. (doi:10.1016/S0893-6080(00)00062-9)
- Hsieh, L., Gandour, J., Wong, D. & Hutchins, G. D. 2001 Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.* **76**, 227–252. (doi:10.1006/brln.2000.2382)
- Hugdahl, K., Bronnick, K., Kyllingsbaek, S., Law, I., Gade, A. & Paulson, O. 1999 Brain activation during dichotic presentations of consonant–vowel and musical instrument stimuli: a 15O-PET study. *Neuropsychologia* **37**, 431–440. (doi:10.1016/S0028-3932(98)00101-8)
- 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.
- Jamison, H. L., Watkins, K. E., Bishop, D. V. M. & Matthews, P. M. 2006 Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* **16**, 1266–1275. (doi:10.1093/cercor/bhj068)
- Joanisse, M. F. & Gati, J. S. 2003 Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage* **19**, 64.
- Johnsruide, I. S., Zatorre, R. J., Milner, B. A. & Evans, A. C. 1997 Left-hemisphere specialization for the processing of acoustic transients. *Neuroreport* **8**, 1761–1765. (doi:10.1097/00001756-199705060-00038)
- Johnsruide, I. S., Penhune, V. B. & Zatorre, R. J. 2000 Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* **123**, 155–163. (doi:10.1093/brain/123.1.155)
- Klein, D., Milner, B. A., Zatorre, R. J., Zhao, V. & Nikelski, E. J. 1999 Cerebral organization in bilinguals: a PET study of Chinese–English verb generation. *Neuroreport* **10**, 2841–2846. (doi:10.1097/00001756-199909090-00026)
- Klein, D., Zatorre, R., Milner, B. & Zhao, V. 2001 A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage* **13**, 646–653. (doi:10.1006/nimg.2000.0738)
- Kluender, K. R., Diehl, R. L. & Killeen, P. R. 1987 Japanese quail can learn phonetic categories. *Science* **237**, 1195–1197. (doi:10.1126/science.3629235)
- Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G. & Friederici, A. D. 2002 Bach speaks: a cortical “language-network” serves the processing of music. *NeuroImage* **17**, 956–966. (doi:10.1016/S1053-8119(02)91154-7)
- Krishnan, A., Xu, Y., Gandour, J. T. & Cariani, P. A. 2004 Human frequency-following response: representation of pitch contours in Chinese tones. *Hear. Res.* **189**, 1–12. (doi:10.1016/S0378-5955(03)00402-7)
- Krishnan, A., Xu, Y., Gandour, J. & Cariani, P. 2005 Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* **25**, 161–168. (doi:10.1016/j.cogbrainres.2005.05.004)
- Kuhl, P. K. 2000 A new view of language acquisition. *Proc. Natl Acad. Sci. USA* **97**, 11 850–11 857. (doi:10.1073/pnas.97.22.11850)
- Kuhl, P. K. & Miller, J. D. 1975 Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. *Science* **190**, 69–72. (doi:10.1126/science.1166301)
- Kujala, T., Alho, K. & Näätänen, R. 2000 Cross-modal reorganization of human cortical functions. *Trends Neurosci.* **23**, 115–120. (doi:10.1016/S0166-2236(99)01504-0)
- Li, X., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Lowe, M. & Tong, Y. 2003 Selective attention to lexical tones recruits left dorsal frontoparietal network. *Neuroreport* **14**, 2263–2266. (doi:10.1097/00001756-200312020-00025)
- Lieberman, A. M. & Mattingly, I. G. 1985 The motor theory of speech perception revised. *Cognition* **21**, 1–36. (doi:10.1016/0010-0277(85)90021-6)
- Lieberman, A. M. & Whalen, D. H. 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. R., Spitzer, S. M., Possing, E. T. & Medler, D. A. 2005 Neural substrates of phonemic perception. *Cereb. Cortex* **15**, 1621–1631. (doi:10.1093/cercor/bhi040)
- Liégeois-Chauvel, C., de Graaf, J., Laguitton, V. & Chauvel, P. 1999 Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb. Cortex* **9**, 484–496. (doi:10.1093/cercor/9.5.484)
- MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C. R., Suckling, J., Calvert, G. A. & Brammer, M. J. 2002 Neural systems underlying British sign language and audio-visual English processing in native users. *Brain* **125**, 1583–1593. (doi:10.1093/brain/awf153)
- McIntosh, A. 2000 Towards a network theory of cognition. *Neural Netw.* **13**, 861–870. (doi:10.1016/S0893-6080(00)00059-9)
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G. & von Cramon, D. Y. 2002 FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* **17**, 73–88. (doi:10.1002/hbm.10042)
- Meyer, M., Alter, K. & Friederici, A. 2003 Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. *J. Neurolinguist.* **16**, 277–300. (doi:10.1016/S0911-6044(03)00026-5)
- Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J. & Dooling, R. J. 1976 Discrimination and labelling of noise–buzz sequences with varying noise-lead times: an example of categorical perception. *J. Acoust. Soc. Am.* **60**, 410–417. (doi:10.1121/1.381097)

- Milner, B. A. 1962 Laterality effects in audition. In *Interhemispheric relations and cerebral dominance* (ed. V. Mountcastle), pp. 177–195. Baltimore, MD: Johns Hopkins Press.
- Minagawa-Kawai, Y., Mori, K., Furuya, I., Hayashi, R. & Sato, Y. 2002 Assessing cerebral representations of short and long vowel categories by NIRS. *Neuroreport* **13**, 581–584. (doi:10.1097/00001756-200204160-00009)
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., 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. *et al.* 1997 Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434. (doi:10.1038/385432a0)
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D. & Matthews, P. M. 2003 Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* **13**, 1362–1368. (doi:10.1093/cercor/bhg083)
- Nenonen, S., Shestakova, A., Huottilainen, M. & Naatanen, R. 2003 Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Brain Res. Cogn. Brain Res.* **16**, 492–495. (doi:10.1016/S0926-6410(03)00055-7)
- Patel, A. & Balaban, E. 2001 Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat. Neurosci.* **4**, 839–844. (doi:10.1038/90557)
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. 2002 The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767–776. (doi:10.1016/S0896-6273(02)01060-7)
- Penhune, V. B., Zatorre, R. J. & Evans, A. C. 1998 Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* **10**, 752–765. (doi:10.1162/089892998563149)
- Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E. & Evans, A. C. 1999 Localization of cerebral activity during simple singing. *NeuroReport* **10**, 3979–3984.
- Petersen, S., Fox, P., Posner, M., Mintun, M. & Raichle, M. 1988 Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* **331**, 585–589. (doi:10.1038/331585a0)
- Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D. & Evans, A. C. 2000 Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proc. Natl Acad. Sci. USA* **97**, 13 961–13 966. (doi:10.1073/pnas.97.25.13961)
- Phillips, D. P. & 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. (doi:10.1016/0166-4328(90)90001-U)
- Pisoni, D. B. 1977 Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *J. Acoust. Soc. Am.* **61**, 1352–1361. (doi:10.1121/1.381409)
- Plante, E., Creusere, M. & Sabin, C. 2002 Dissociating sentential prosody from sentence processing: activation interacts with task demands. *NeuroImage* **17**, 401–410. (doi:10.1006/nimg.2002.1182)
- Poeppl, D. 2003 The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* **41**, 245–255. (doi:10.1016/S0167-6393(02)00107-3)
- Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D. 1981 Speech perception without traditional speech cues. *Science* **212**, 947–949. (doi:10.1126/science.7233191)
- Robin, D. A., Tranel, D. & Damasio, H. 1990 Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang.* **39**, 539–555. (doi:10.1016/0093-934X(90)90161-9)
- Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G. & Hallett, M. 1996 Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **380**, 526–528. (doi:10.1038/380526a0)
- Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C. & Chen, H. C. 2005 Brain responses to segmentally and tonally induced semantic violations in Cantonese. *J. Cogn. Neurosci.* **17**, 1–12. (doi:10.1162/0898929052880057)
- Schönwiesner, M., Rubsamen, R. & von Cramon, D. Y. 2005 Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* **22**, 1521–1528. (doi:10.1111/j.1460-9568.2005.04315.x)
- Scott, S. K. & Johnsrude, I. S. 2003 The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100. (doi:10.1016/S0166-2236(02)00037-1)
- Scott, S. K. & Wise, R. S. J. 2004 The functional neuroanatomy of prelexical processing in speech perception. *Cognition* **92**, 13–45. (doi:10.1016/j.cognition.2002.12.002)
- Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. S. 2000 Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406. (doi:10.1093/brain/123.12.2400)
- Shaywitz, B. A. *et al.* 2001 The functional neural architecture of components of attention in language-processing tasks. *NeuroImage* **13**, 601–612. (doi:10.1006/nimg.2000.0726)
- Shtyrov, Y., Pihko, E. & Pulvermüller, F. 2005 Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *NeuroImage* **27**, 37–47. (doi:10.1016/j.neuroimage.2005.02.003)
- Sidtis, J. J. & Volpe, B. T. 1988 Selective loss of complex-pitch or speech discrimination after unilateral lesion. *Brain Lang.* **34**, 235–245. (doi:10.1016/0093-934X(88)90135-6)
- Sittiprapaporn, W., Chindaduangratn, C., Tervaniemi, M. & Khotchabhakdi, N. 2003 Preattentive processing of lexical tone perception by the human brain as indexed by the mismatch negativity paradigm. *Ann. NY Acad. Sci.* **999**, 199–203. (doi:10.1196/annals.1284.029)
- Swisher, L. & Hirsh, I. J. 1972 Brain damage and the ordering of two temporally successive stimuli. *Neuropsychologia* **10**, 137–152. (doi:10.1016/0028-3932(72)90053-X)
- Tallal, P., Miller, S. & Fitch, R. 1993 Neurobiological basis of speech: a case for the preeminence of temporal processing. *Ann. NY Acad. Sci.* **682**, 27–47. (doi:10.1111/j.1749-6632.1993.tb22957.x)
- Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M. & Merzenich, M. M. 1996 Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* **271**, 81–84. (doi:10.1126/science.271.5245.81)
- Tervaniemi, M., Medvedev, S., Alho, K., Pakhomov, S., Roudas, M., van Zuijen, T. & Näätänen, R. 2000 Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* **10**, 74–79. (doi:10.1002/(SICI)1097-0193(200006)10:2<74::AID-HBM30>3.0.CO;2-2)
- Tong, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Xu, Y., Li, X. & Lowe, M. 2005 Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage* **28**, 417–428. (doi:10.1016/j.neuroimage.2005.06.002)
- Valaki, C. E., Maestu, F., Simos, P. G., Zhang, W., Fernandez, A., Amo, C. M., Ortiz, T. M. & Papanicolaou, A. C. 2004 Cortical organization for receptive language functions in Chinese, English, and



- Spanish: a cross-linguistic MEG study. *Neuropsychologia* **42**, 967–979. (doi:10.1016/j.neuropsychologia.2003.11.019)
- von Steinbüchel, N. 1998 Temporal ranges of central nervous processing: clinical evidence. *Exp. Brain Res.* **123**, 220–233. (doi:10.1007/s002210050564)
- Vouloumanos, A., Kiehl, K. A., Werker, J. F. & Liddle, P. F. 2001 Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* **13**, 994–1005. (doi:10.1162/089892901753165890)
- Wang, Y., Sereno, J. A., Jongman, A. & Hirsch, J. 2003 FMRI evidence for cortical modification during learning of Mandarin lexical tone. *J. Cogn. Neurosci.* **15**, 1019–1027. (doi:10.1162/089892903770007407)
- Warrier, C. M. & Zatorre, R. J. 2004 Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain* **127**, 1616–1625. (doi:10.1093/brain/awh183)
- Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C., Cohen, L., Hallett, M. & Rauschecker, J. 2000 A positron emission tomographic study of auditory localization in the congenitally blind. *J. Neurosci.* **20**, 2664–2672.
- Whalen, D. H. & Liberman, A. L. 1987 Speech perception takes precedence over nonspeech perception. *Science* **237**, 169–171. (doi:10.1126/science.3603014)
- Wong, P. 2002 Hemispheric specialization of linguistic pitch patterns. *Brain Res. Bull.* **59**, 83–95.
- Wong, P. C., Parsons, L. M., Martinez, M. & Diehl, R. L. 2004 The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J. Neurosci.* **24**, 9153–9160. (doi:10.1523/JNEUROSCI.2225-04.2004)
- Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., Li, X. & Lowe, M. 2006 Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum. Brain Mapp.* **27**, 173–183. (doi:10.1002/hbm.20176)
- Yip, M. 2003 *Tone*. New York, NY: Cambridge University Press.
- Zaehle, T., Wustenberg, T., Meyer, M. & Jancke, L. 2004 Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *Eur. J. Neurosci.* **20**, 2447–2456. (doi:10.1111/j.1460-9568.2004.03687.x)
- Zatorre, R. J. 1983 Category-boundary effects and speeded sorting with a harmonic musical interval continuum. *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 739–752. (doi:10.1037/0096-1523.9.5.739)
- Zatorre, R. J. 1988 Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* **84**, 566–572. (doi:10.1121/1.396834)
- Zatorre, R. J. & Belin, P. 2001 Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953. (doi:10.1093/cercor/11.10.946)
- Zatorre, R. J. & Halpern, A. R. 1993 Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia* **31**, 221–232. (doi:10.1016/0028-3932(93)90086-F)
- Zatorre, R. J. & Samson, S. 1991 Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain* **114**, 2403–2417. (doi:10.1093/brain/114.6.2403)
- Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. 1992 Lateralization of phonetic and pitch processing in speech perception. *Science* **256**, 846–849. (doi:10.1126/science.1589767)
- Zatorre, R. J., Evans, A. C. & Meyer, E. 1994 Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* **14**, 1908–1919.
- Zatorre, R. J., Meyer, E., Gjedde, A. & Evans, A. C. 1996 PET studies of phonetic processing of speech: review, replication, and re-analysis. *Cereb. Cortex* **6**, 21–30. (doi:10.1093/cercor/6.1.21)
- Zatorre, R. J., Mondor, T. A. & Evans, A. C. 1999 Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage* **10**, 544–554. (doi:10.1006/nimg.1999.0491)
- Zatorre, R. J., Belin, P. & Penhune, V. B. 2002 Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46. (doi:10.1016/S1364-6613(00)01816-7)