

# Cortical asymmetries in speech perception: what's wrong, what's right and what's left?

Carolyn McGettigan and Sophie K. Scott

Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK

**Over the past 30 years hemispheric asymmetries in speech perception have been construed within a domain-general framework, according to which preferential processing of speech is due to left-lateralized, non-linguistic acoustic sensitivities. A prominent version of this argument holds that the left temporal lobe selectively processes rapid/temporal information in sound. Acoustically, this is a poor characterization of speech and there has been little empirical support for a left-hemisphere selectivity for these cues. In sharp contrast, the right temporal lobe is demonstrably sensitive to specific acoustic properties. We suggest that acoustic accounts of speech sensitivities need to be informed by the nature of the speech signal and that a simple domain-general vs. domain-specific dichotomy may be incorrect.**

## Explaining the imbalance

It is generally well agreed that, for most people, both speech production and speech perception, as it serves language, are functions of the left hemisphere [1,2]. The vast majority of patients with aphasia have a left-hemisphere lesion, and a left dominance for speech and language has also been reported in intact brains using dichotic listening, transcranial magnetic stimulation (TMS) and functional neuroimaging (fMRI). One influential body of research over the past two decades has concentrated on the hypothesis that this left dominance might emerge from a sensitivity to non-linguistic acoustic information which happens to be critical for speech. In approaches strongly influenced by hypotheses about the role of temporal information in speech (Box 1), resulting theories proposed that the left hemisphere would show a preference for processing of rapid temporal information in sound. For example, Robert Zatorre and colleagues specified a model in which left auditory fields were sensitive to temporal structure and the right to spectral structure, by direct analogy with the construction of wide-band and narrow-band spectrograms [3]. At approximately the same time, David Poeppel advanced a theory informed by the patterns of oscillations studied in electrophysiological approaches (Box 2). The Asymmetric Sampling in Time hypothesis [4] proposed that the left temporal lobe preferentially samples information over short time windows comparable to the duration of phonemes, and the right over longer time windows

comparable to the length of syllables and intonation profiles. These theories have been widely used as explanatory frameworks over the past decade.

## The nature of speech

It is important to consider whether these neural models make reasonable assumptions about the nature of the information conveyed in speech (Figure 1). The approaches of Zatorre and Poeppel both clearly associate left auditory fields with a sensitivity to temporal or rapid information. However, this may be an inaccurate characterization of speech:

- Not all the information in speech exists over short time scales. Plosives, which are frequently the only consonants studied (possibly because they can be neatly fitted into a matrix of place/manner/presence or absence of voicing cues, unlike many other speech sounds), do include very rapidly evolving acoustic structure. However, fricatives, affricates, nasals, liquids and vowels are often considerably longer in duration than the (40 ms) window specified by Poeppel.
- As any sound, by its very nature, can only exist over time, 'temporal' is not a very well specified term. There is no non-temporal information available in sound. In turn, 'temporal' has been used in a variety of senses; Efron [5] used the concept of time to point out that phonemes need to be heard in the right order to result in meaningful speech, whereas Tallal and Piercy [6] expressly considered the processing of temporal order judgments to reflect similar processes to those necessary for the identification of phonemes in the incoming speech stream. More recently, 'temporal' has been used to refer to the amount of smoothing applied to the amplitude envelope of speech [7].
- Phonetic information in speech is not expressed solely over short time scales and occurs over a supra-segmental level. Consecutive articulatory manoeuvres are not independent and this co-articulation is reflected in the acoustics. This process can have effects over adjacent speech sounds as the articulation of one sound anticipates the next (e.g. the 's' at the start of 'sue' is very different from the 's' at the start of 'see' [8]). These anticipatory effects can also take place over the whole syllable: the 'l' sound at the start of 'led' is different from that at the start of 'let' because of the voicing of the final syllable. Listeners are sensitive to these cues [9].

Corresponding author: Scott, S.K. (sophie.scott@ucl.ac.uk).

### Box 1. a journey in time and space

Nearly 50 years ago, Robert Efron [5] suggested that the processing of temporal information and speech may share some resources, showing that patients with aphasia were impaired at overt temporal order judgments, needing longer intervals between two stimuli before they could determine which came first. Efron's finding inspired a more general interest in how speech perception might be supported by non-linguistic processes and raised the possibility that hemispheric specializations in speech perception could be accounted for in a domain-general manner. It was striking that the temporal ordering problems exhibited by the patients were notable from quite long inter-stimulus intervals (ISIs – around 400 ms), which in terms of speech perception are on the order of syllables, rather than phonemes.

Tallal and Piercy [6] investigated how children with aphasia (who today would likely be diagnosed with specific language impairment) performed on auditory temporal order judgments. These children, like the adult patients with aphasia, needed longer ISIs than controls to be able to discern which sound came first. As in Efron's study, the children with 'aphasia' began to show problems at relatively long ISIs (around 300 ms). These findings have been widely interpreted as demonstrations of problems with rapid temporal sequencing.

- No single acoustic cue underlies the way that information is expressed in speech, and any one 'phonetic' contrast is underpinned by a variety of acoustic features. In British English, the phonetic difference between /apa/ and /aba/ concerns only the voicing of the medial /p/ or /b/ phones; however, there are more than 16 separable acoustic features which differ between the two plosives, including the length of the pre-stop vowel being longer in /aba/ and the aspiration on the release of /p/ being more pronounced than for /b/ [10].
- All sounds can be characterized in terms of their spectral and amplitude modulation profiles – in speech, for example, most of the energy is carried in low spectral and amplitude modulations. Speech intelligibility can be preserved when the spectral and amplitude modulations are quite coarse (e.g. listeners can understand speech vocoded to 6 broad spectral channels and considerable smoothing of the amplitude envelope); however, neither kind of modulation in isolation can yield an intelligible percept. Given that speech perception requires spectral

modulations [11], it is unclear why spectral detail should be processed preferentially in the right hemisphere (e.g. [3,7]). In some cases, 'spectral' is used to mean 'pitch' [3,12], although these two terms are in fact not synonymous.

Finally, a serious issue for both the approaches of Poeppel and Zatorre is the identification of the phoneme as the fundamental unit of perceptual information in speech. That is, the assumption that access to phoneme representations is the cardinal aspect of speech perception, and hence that theories need to account for a left-dominant sensitivity to phenomena which might underlie phoneme perception. There is considerable evidence that this assumption is incorrect [13–18]. Indeed, the growing evidence for the importance of suprasegmental structures like onset-rime structure, or syllables, would surely call for a temporal account that applies a wider time window on the speech signal, on the order of hundreds, rather than tens, of milliseconds.

### Box 2. Neural oscillations and temporal primitives

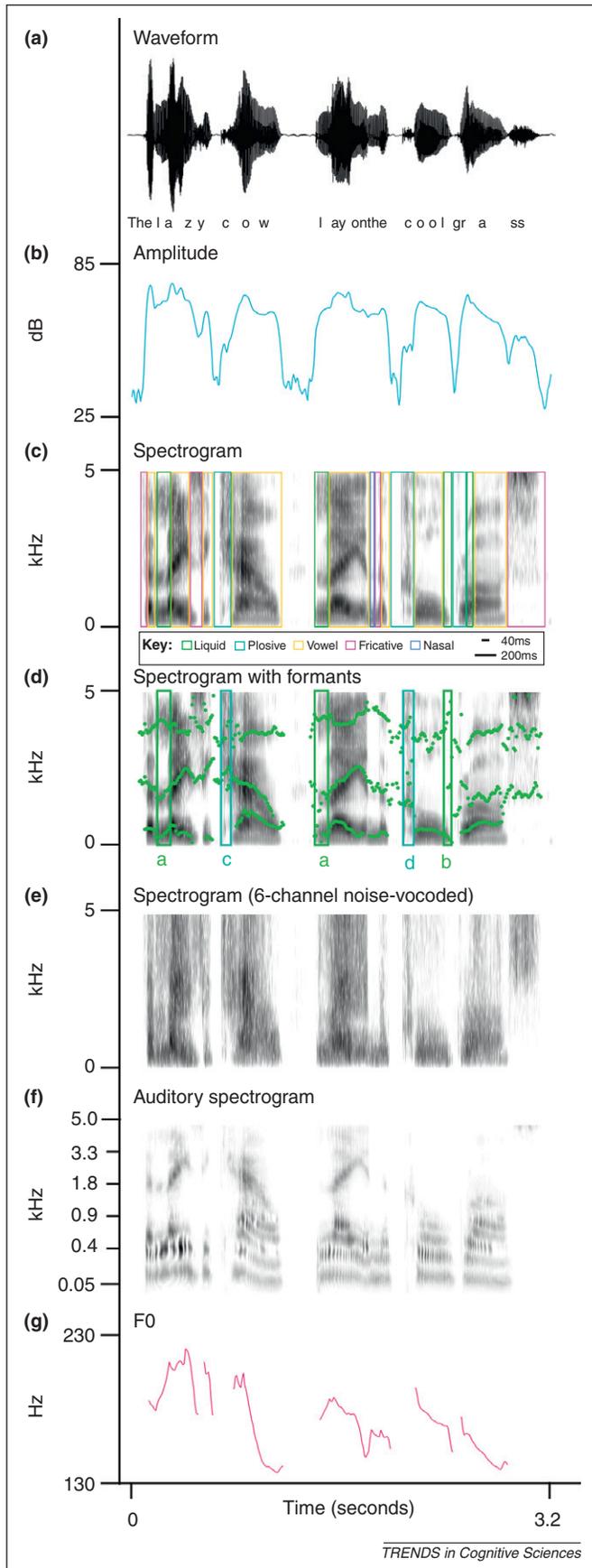
The Asymmetric Sampling in Time hypothesis [4] predicts that ongoing oscillatory activity in different frequency bands forms computational primitives in the brain. Theta range (4-8 Hz) activity should be maximally sensitive to information at the rate of syllables in speech, whereas neuronal populations operating in the low gamma range (~40 Hz) should be sensitive to the rapid temporal changes assumed to be central to phonemic information. The theory proposes a predominance of gamma populations in the left auditory cortex and of theta on the right [66]. The correlation between ongoing electroencephalography (EEG) and BOLD activity at rest supports this view for the right hemisphere [66].

Are these different oscillations integral to the processing of speech? Phase-locked oscillatory theta activity can discriminate between three heard sentences but not when stimulus intelligibility is degraded [67]. Theta rhythms track incoming information in a stimulus-specific manner and this tracking reflects more than slavish responsivity to the sentence acoustics [68]. Phase-locked delta-theta (2-7 Hz) responses support the coordination of incoming audio and visual streams [69]. When listeners were attending to the speech of one of two simultaneous talkers, significant sentence discrimination was possible using phase-locked activity in the 4-8 Hz (theta) range, which was modulated by selective attention [7]. Moreover, in partial support

of the AST model, these theta effects were right-dominant. The complementary gamma activity proposed by the AST model is relatively lacking in these studies (but see [8,9]). Notably, speech perception is not restricted to gamma and theta frequencies. Selective attention to speech is associated with lateralized alpha (8-13 Hz) activity [70], correlated with enhanced theta activity in auditory regions. Neural responses to degraded words show an alpha suppression, some 0.5 seconds after word onset, sensitive to both spectral detail and comprehension ratings [71].

The inter-relationship of measures available from electrophysiological recordings is complex. Thus, the differences in functional significance between induced changes in power, evoked responses, phase-resetting and coherence are not yet fully understood [72]. Moreover, higher-frequency modulations can become 'nested' in lower frequency ranges [73,74].

Combined EEG and fMRI has shown that during learning a significant correlation between a beta power decrease and the BOLD response in left inferior frontal gyrus is specific to the encoding of later-remembered items [75]. A similar method could be used to identify the spatial location of task-critical oscillatory activity in response to heard speech and asymmetries might be seen in the network dynamics connecting speech-sensitive nodes in the two hemispheres.



**Figure 1.** What's in a sound? Speech is an immensely complex signal. **(a)** The waveform of a spoken sentence, plotted as an amplitude/time profile (an oscillogram). **(b)** The amplitude envelope of the sentences, in which the oscillogram is rectified and smoothed. **(c)** A spectrogram of the sentence, in which the frequency components of the sound are represented on the ordinate,

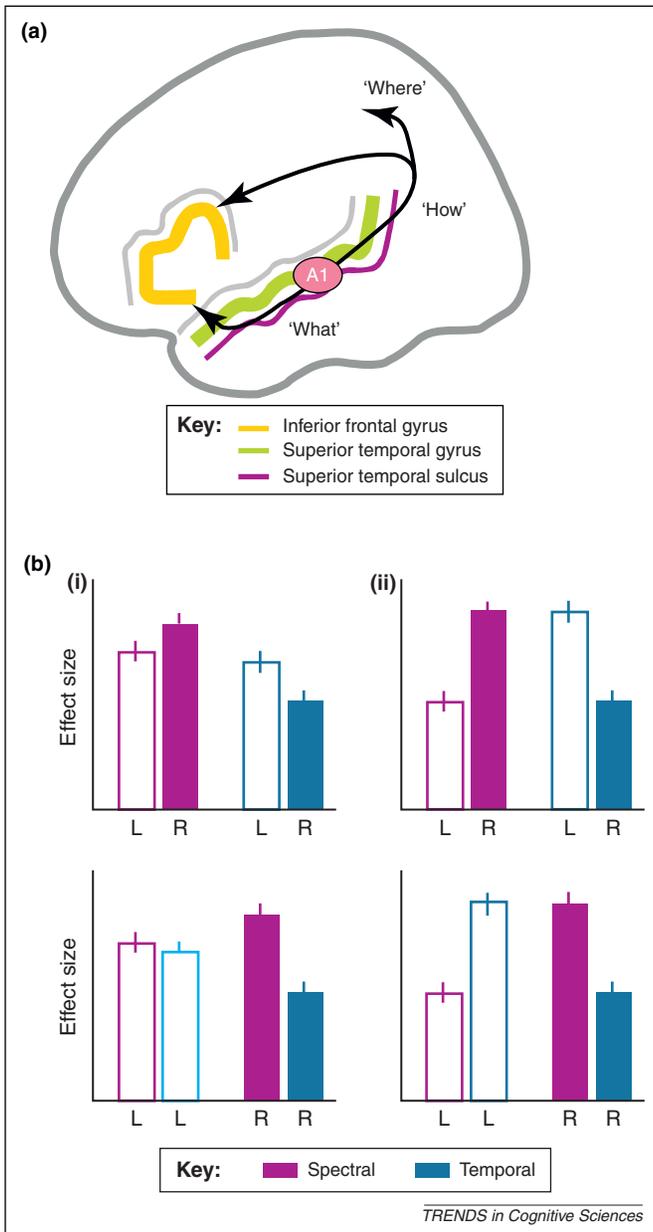
### Hemispheric asymmetries in sound processing

Thus, there are *a priori* reasons to suggest that a simple left–right dichotomy of acoustic processing may not account for the asymmetric processing of speech. In line with these reservations, it has proved difficult to show unambiguous evidence for a left-lateralized dominance for ‘temporal’ processing of sounds, regardless of the size of the sampling window. Left-lateralized neural responses to fast versus slow formant transitions were reported by Johnsrude and colleagues [19], but these lay in the left inferior frontal cortex (see also [20–22]) rather than the early auditory fields in temporal cortex associated with hearing and speech perception (Figure 2). Using elegant speech-like stimuli that contained either fast and slow formant transitions without being intelligible (English) speech, Belin and colleagues [23] showed that the left superior temporal gyrus (STG), extending into primary auditory cortex (PAC), was sensitive to both fast and slow transitions and that the right STG (extending into PAC) was responsive to slow transitions only. This study is widely interpreted as showing a left dominance for rapid transitions but was actually the first study to show that it is perhaps the right temporal lobe that responds selectively to acoustic manipulations in sounds. A caveat, however, is that the duration of the stimuli co-varied with the rapid/slow transitions, and so it may be that the right temporal lobe simply prefers longer sounds.

This pattern of right hemisphere preference for longer sounds, and sounds with pitch variation, has since been observed in a wide variety of studies [3,12,23,24]. A dominant trend has been to interpret such data as showing a left hemisphere preference for ‘temporal’ or ‘rapid’ temporal processing [25], although none of these studies actually shows a selective sensitivity (i.e. greater activation) in the left hemisphere for rapid or short temporal cues. A claim for a selective response to ‘temporal’ cues on the left has even been made when the neural response in the left superior temporal sulcus is greater to increasing ‘spectral’ detail than to increasing ‘temporal’ detail [7]. Although some studies have found differences in the correlation of the blood oxygenation level-dependent (BOLD) response in the left and right auditory fields with spectral and temporal cues [24], the lack of a direct comparison of the effect sizes of the two parameters makes these results difficult to interpret (Figure 2).

Some studies have reported a left-lateralized or left-dominant response to temporal characteristics of sounds. Zatorre and Belin [3] and Jamieson and colleagues [26]

time on the abscissa, and intensity is expressed by darker colours. This enables the complex spectral structure (formants) in speech to be visualized, reflecting the changes in spectral prominences associated with different movements and positions of the articulators. Different kinds of phonemes are indicated in coloured boxes: note that most of these are longer than a 40-ms time window. **(d)** The same spectrogram, now with the formants specifically being tracked over time. Note that the different // sounds have very different spectral structures (‘clear’ at word onsets (a) and ‘dark’ at word offsets (b)), and that the two /k/ sounds (at the start of ‘cow’ (c) and ‘cool’ (d)) are also different. These are due to phonotactic constraints and co-articulation, respectively. **(e)** A spectrogram of the same sentence noise vocoded to 6 channels: this speech can still be understood although it only contains coarse spectral and amplitude structure. **(f)** An auditory spectrogram, which more accurately reflects the way that frequency information is represented in the human auditory system: note that this greatly magnifies the detail at lower spectral frequencies. **(g)** The fundamental frequency (F0) of the sentence (i.e. the shape of the intonation profile of the sentence).



**Figure 2.** Locations and configurations of hemispheric asymmetries. (a) There are several different distinct auditory, heteromodal and prefrontal regions which are activated in speech and sound processing. Generally, the field is openminded about where (PAC, STG, STS, IFG) hemispheric asymmetries might be identified, although different interpretations might be made of the consequences of asymmetries in these different cortical areas. Thus asymmetries in primary auditory cortex might lead to an assumption that speech (and sounds that share its hypothesized acoustic features) is selectively processed from its earliest encoded entry into the cortex. Asymmetries in the superior temporal gyrus might point to a somewhat later acoustic selectivity process, whereas asymmetric responses in the STS and IFG could arguably be related to non-acoustic sensitivities. Furthermore, the rostral/caudal orientation of anatomically and functionally distinct streams of auditory processing may be important in interpreting the results of such functional imaging studies. (b) The bar charts in the lower section use hypothetical data to illustrate how different ways of plotting effect sizes can lead to varying interpretations of the data. The plots on the right (i) reflect potentially valid hemispheric asymmetries according to the hypothesis of Zatorre and colleagues (i.e. the left responds selectively to temporal information and the right prefers spectral properties), whereas the plots on the left (ii) show effects driven by an acoustic selectivity in the right hemisphere response only. The responses of the left and the right hemisphere are illustrated using clear and filled bars, respectively, with the only difference between the upper and lower panels being the order in which the bars are arranged. Note that the similarities between the two upper panels disappears when the results are plotted by hemisphere, rather than by condition.

showed a dominant left PAC response that correlated with decreasing interstimulus intervals and stimulus durations across pure tones of two different pitches (separated by one octave). Zaehle and colleagues [27] compared the neural responses to noise stimuli containing gaps (where the gaps were either 8 or 32 ms in duration, after a leading element of 5 or 7 ms of noise) with responses to tones, in the context of a delayed match to sample task. This contrast revealed a unilateral response in left medial primary auditory cortex (encroaching on planum temporale and insular cortex) to the noise stimuli with gaps. While all three studies converge on similar conclusions with regard to a left-hemisphere specialization for temporal processing, it is also the case that there are other acoustic factors that need to be considered. All three studies conflate 'temporal' manipulations with changes in the number of onsets of acoustic stimulation (including unwanted onset phenomena, such as spectral splatter). For example, when the rate of change between two tones is low [3,26], there are fewer perceived onset events associated with the change from one pitch to another. When the rate of change is high, there are many more onsets. Furthermore, these studies [3,26] included tone repetition times and durations so brief (around 21 ms) that the sequences could be perceived as two separate auditory streams [28]. Consequently, it is not easy to determine whether the left-lateralized responses arise due to the 'temporal' nature of the manipulations or the fact that the manipulations led to the formation of multiple auditory onsets.

The concept of a left auditory preference for (rapid) temporal processing is certainly appealing, especially if one favours a domain-general rather than a domain-specific argument about the systems underlying the left-lateralized nature of speech [25]. However, approaching a theory of hemispheric asymmetries by positing complementary hemispheric networks might in fact impose a symmetry or balance to the brain that simply is not there. In other words, offering an account that affords a particular acoustic preference to each hemisphere might lead to overlooking a true asymmetry, in which the right hemisphere shows genuine preferences for certain properties of the acoustic signal, whereas the left does not.

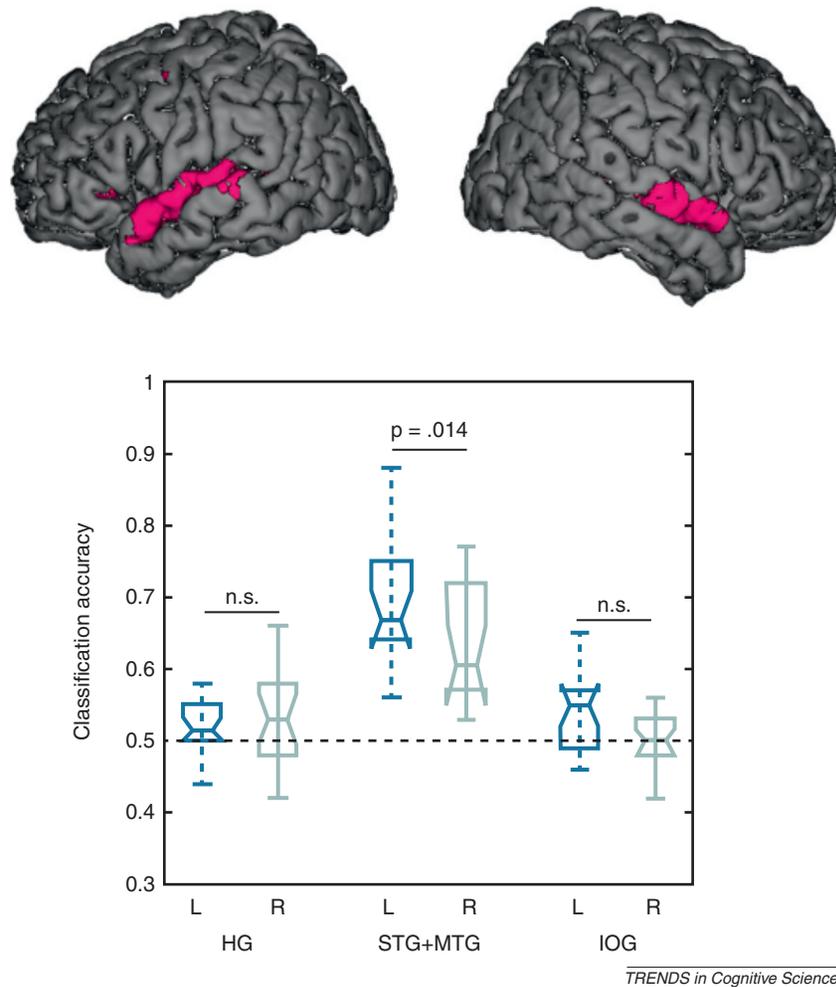
### Hemispheric asymmetries in speech processing

Although it is generally assumed that speech perception is left-dominant in the human brain, methodologically it has not been simple to identify a satisfactory means of hemispheric comparisons using the standard univariate analysis of the signal obtained in PET or fMRI (Box 3). In contrast to the studies on spectral/temporal information or time windows, straightforward comparisons of conditions (e.g. speech > acoustically-matched control stimuli) has shown clearer evidence of left-dominant responses to speech. This preference has been shown for sentences [29–33], words [34], syllables [35], isolated phonemes [36] and phonologically relevant acoustic cues [37], as well as syntactic [38–40] and semantic information [41,42]. This dominance is typically seen both in the extent of activation and in the size of the effect (although note the limitations of interpretation for these parameters discussed in Box 3), and is most easily interpreted when right-sided activation

### Box 3. You say lateral, I say (bi)lateral

A problem in the investigation of functional hemispheric asymmetries is how to offer sufficient proof that an effect is present or not. Laterality indices have been calculated using the difference in the peak statistical height and spatial extent of activations in the two hemispheres [7,26]. However, counting voxels makes unwarranted assumptions about the way that thresholding the data affects the extent of activation. In other cases, authors have 'flipped' images to facilitate subtractive comparisons of the left and right hemispheres (e.g. [27,76]) but this does not take account of the anatomical asymmetries present between the hemispheres. Further difficulties arise in the selection of regions-of-interest (ROI) for targeted extraction of local data in the left and right temporal lobes. Selection of ROIs based on functional data, or using the coordinates of peak activations from the same or previous studies, potentially introduces problems of non-independence and arbitrariness to the analyses. Moreover, the sensitivity of the mean BOLD signal to subtle differences between conditions is compromised in large ROIs [77], which is an important consideration when the effects at hand are hypothesized to be subtle [4].

A recent study [31] introduced multivariate pattern analysis of unsmoothed fMRI data in anatomically-defined ROIs as a means of directly compare acoustic and linguistic processing in the left and right temporal lobes. A linear support vector machine (SVM) was used to train and validate models in the classification of pairs of speech-derived conditions that varied in their acoustic properties and intelligibility (see main text and Figure 1). This method presents several advantages over previous approaches. First, the data were not subject to thresholding and its effects on apparent activation extent. Second, the ROIs were anatomically defined for each subject through parcellation of the cortical surface, allowing for the cross-hemispheric comparison of processing within task-relevant (STG, Heschl's Gyrus) and control (inferior occipital gyrus) structures. Moreover, as SVMs are relatively robust to large numbers of voxels [78], this approach avoided the arbitrariness of small ROI selection without compromising the ability to detect effects.



**Figure 1.** Determining asymmetries with multivariate pattern analysis. The rendered cortical surface images show activation resulting from a contrast of intelligible speech with acoustically-matched control sounds. The boxplot illustrates multivariate classification of responses to intelligible and unintelligible items in left and right Heschl's gyrus (HG), superior temporal cortex (STG+MTG) and inferior occipital gyrus (IOG). MTG, middle temporal gyrus. Adapted and redrawn from data presented in [31].

is not found. Arguments that the domain-general processing of sound might operate in a relatively subtle way are at odds with the strong left-dominant responses to different kinds of linguistic information in speech, and with the commonly seen preferential response to sounds with pitch variation on the right.

### Indexical cues in speech – the right for a voice?

Speech contains multiple cues in addition to the linguistic signal, and so cortical responses to undistorted speech may also include some of the speaker indexical information that has been shown to be associated with the right temporal lobe [43–50]. It is possible that aspects of this rightward

preference for speaker-related information may indeed be associated with an acoustic preference, and the right temporal lobe responds more than the left to sounds that are longer, have pitch variation or musical structure [23,50–52]. One candidate explanation, in line with work on the ‘temporal voice areas’ in the right dorsolateral temporal lobe [43], is that the right hemisphere is sensitive to speaker-related acoustic information in speech, above and beyond the linguistic signal (see also [53] for a discussion of how this asymmetry might bear on the processing of vocal emotion processing). It is certainly striking that the non-linguistic specialization which can be demonstrated in auditory cortex is associated with non-linguistic processing of sounds in the right hemisphere, rather than with speech or the left hemisphere.

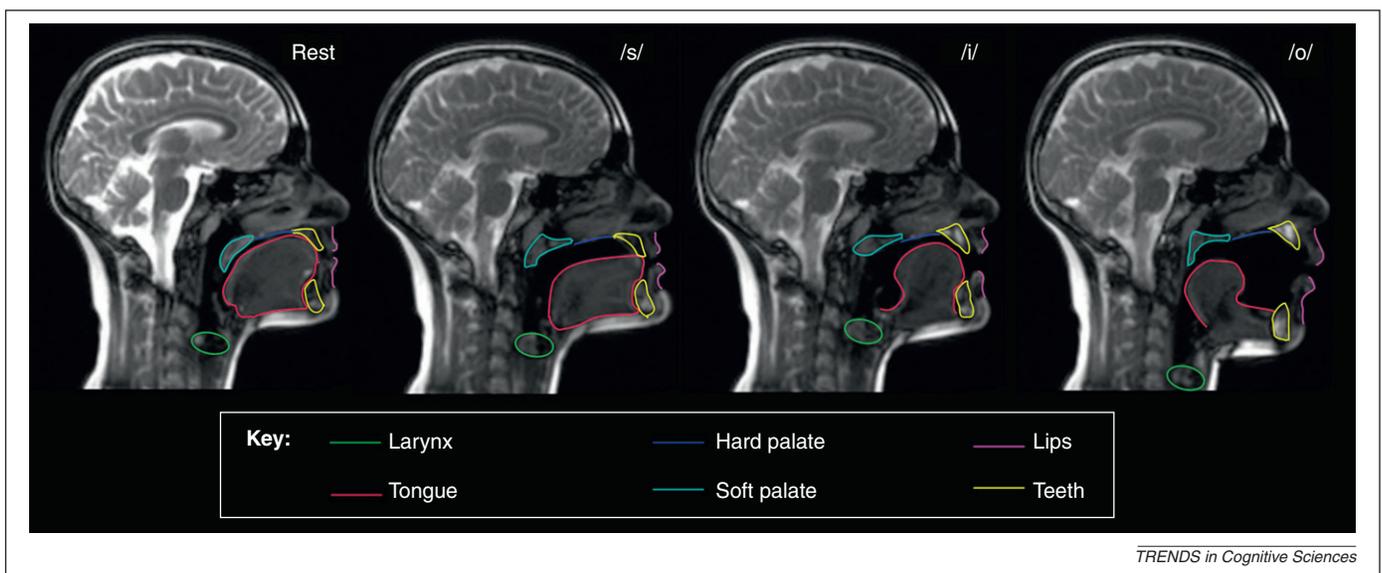
### Information and modulations in speech

If rapidity/temporal information is not key to speech perception (and, as discussed above, this is at best a weak characterization of the nature of speech acoustics), then are there other acoustic cues to which the left temporal lobe might be especially sensitive? Two recent studies addressed this question, using PET [31,54] and fMRI [31] to measure the neural response to amplitude and spectral modulations derived from actual speech signals. The experimental stimuli were constructed such that these different types of modulation could be investigated separately and in combination. A further manipulation allowed for the comparison of conditions where the combination of amplitude and spectral detail led to an intelligible speech percept or to an unintelligible sequence (when the amplitude and spectral modulations came from different sentences). The results of these studies showed strongly bilateral responses to amplitude modulations, spectral modulations and both in combination (when unintelligible). In the PET study, there was a numerical indication that the right STG responded additively to the combined (unintelligible) modulations, whereas the left hemisphere did not. In both studies, the contrast of the intelligible and unintelligible conditions was associated

with left-dominant effects and in the fMRI study this was confirmed statistically with direct comparison of left and right superior temporal regions using multivariate pattern classification (see Box 3). The unintelligible stimuli, which had comparable amplitude and spectral modulations to the intelligible condition, did not lead to left-lateralized effects, suggesting that the presence of the amplitude and spectral modulations is not sufficient to yield a speech-selective response profile in the brain.

### Beyond ‘temporal’ – the instruments of speech

Does this mean that there can be no acoustic factors that might underlie the left-dominant responses to speech? Almost certainly not. However, those acoustic factors will not be simple and they will be dictated by the nature of speech and the way it is produced [55]. For example, the studies described above [31,54] explored the neural responses to spectral and amplitude modulations that are sufficient for speech comprehension, revealing a left-dominant effect only when those combinations could be understood. However, for the unintelligible combinations in those studies, it is possible that the manipulations altered the structure of the syllables, for example the onset/rhyme characteristics of the sequences (which are associated with amplitude onsets in mid-band spectral energy). Recent developments in speech science have demonstrated that multiple correlated cues are used by listeners when processing speech and these cues are tethered as concurrent consequences of mechanics and dynamics of articulation [56]. It is possible that brain activations that are selective for speech may reflect experience with these patterns as they occur within the sound structures of a language. Figure 3 shows examples of the configurations of human articulators during different kinds of speech sound production. Although the tongue is positioned with great flexibility and range to make speech sounds, the movements of the articulators are constrained by the anatomy, such that certain configurational patterns are difficult or impossible – the speech sound ‘s’ is not possible with the



TRENDS in Cognitive Sciences

**Figure 3.** Configurations of the vocal tract. This figure shows images from a real-time MRI sequence during which an adult speaker of British English produced continuous speech. The leftmost panel shows the articulators at rest. The next panels show the configurations associated with the speech sounds, ‘s’, ‘eee’ and ‘aw’, respectively.

#### Box 4. Outstanding questions

- To what extent is speech perception left-lateralized purely due to the left lateralized nature of language and what factors underlie this?
- Do these left-hemisphere preferences for speech and language interact with left/right attentional systems and if so, how?
- What is the functional relationship between the dominant hemisphere and handedness?

jaw configuration of 'o'. Furthermore, speech requires the controlled flow of air (voiced or not) and the movements of the articulators effect changes in the emphasis of the air flow by constriction (which is what is generally meant by place of articulation). This means that many possible configurations of the vocal tract are never used in speech: there are no nasalized /s/ sounds and no glottal sonorant sounds [57].

A domain-general approach to speech perception will also require a more complex, dynamic model of the perceptual processes involved. The perceptual processing of sounds, even after only a brief exposure, is strongly affected by learnt expectations [58,59]. It is difficult to reconcile models which rely on basic acoustic factors (and which are implicitly feed-forward in their approach) with the ways that context and expectations influence perception (which require a more overt incorporation of bi-direction processes (e.g. predictive coding [60,61]).

#### Concluding remarks: what's right and what's left?

Need we abandon the pleasing simplicity of dimensions of auditory sensitivities, which lead to different patterns of processing in the left and right temporal lobes? No, in that the right temporal lobe is strongly selective for certain aspects of the signal; yes, in that the left temporal lobe is not. The complex pattern emerging suggests domain-general mechanisms at play in the right temporal lobe, and more domain-specific mechanisms in the left (see Box 4). Furthermore, if there are speech-selective acoustic sensitivities in the left hemisphere, these acoustic properties are highly unlikely to be dissociable from the articulatory systems and mechanisms that produce the sounds of speech.

#### Acknowledgements

C.M. and S.K.S. are funded by Wellcome Trust Grant WT074414MA awarded to S.K.S. We would like to thank Stuart Rosen, Keith Kluender and Jonas Obleser for very helpful discussions on this topic.

#### References

- 1 Markus, H.S. and Boland, M. (1992) Cognitive activity monitored by noninvasive measurement of cerebral blood-flow velocity and its application to the investigation of cerebral-dominance. *Cortex* 28, 575–581
- 2 Knecht, S. et al. (2000) Handedness and hemispheric language dominance in healthy humans. *Brain* 123, 2512–2518
- 3 Zatorre, R.J. and Belin, P. (2001) Spectral and temporal processing in human auditory cortex. *Cereb. cortex* 11, 946–953
- 4 Poeppel, D. (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Comm.* 41, 245–255
- 5 Efron, R. (1963) Temporal perception, aphasia and *deja vu*. *Brain* 86, 403–424
- 6 Tallal, P. and Piercy, M. (1973) Defects of nonverbal auditory-perception in children with developmental aphasia. *Nature* 241, 468–469
- 7 Obleser, J. et al. (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8123
- 8 Bailey, P.J. and Summerfield, Q. (1980) Information in speech: observations on the perception of [s]-stop clusters. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 536–563
- 9 Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon.* 31, 373–405
- 10 Lisker, L. (1977) Rapid versus ravid - catalog of acoustic features that may cue distinction. *J. Acoust. Soc. Am.* 62, S77–S78
- 11 Stilp, C.E. et al. (2010) Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *J. Acoust. Soc. Am.* 128, 2112–2126
- 12 Boemio, A. et al. (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395
- 13 McGettigan, C. et al. (2011) Neural Correlates of Sublexical Processing in Phonological Working Memory. *J. Cogn. Neurosci.* 23, 961–977
- 14 Vaden, K.I., Jr et al. (2011) Sublexical properties of spoken words modulate activity in broca's area but not superior temporal cortex: implications for models of speech recognition. *J. Cogn. Neurosci.* 23, 2665–2674
- 15 Boucher, V.J. (1994) Alphabet-related biases in psycholinguistic inquiries - considerations for direct theories of speech production and perception. *J. Phon.* 22, 1–18
- 16 Port, R. (2007) How are words stored in memory? Beyond phones and phonemes. *New Ideas Psychol.* 25, 143–170
- 17 Port, R.F. (2010) Rich memory and distributed phonology. *Lang. Sci.* 32, 43–55
- 18 Port, R.F. (2010) Language as a Social institution: why phonemes and words do not live in the brain. *Ecol. Psychol.* 22, 304–326
- 19 Johnsrude, I.S. et al. (1997) Left-hemisphere specialization for the processing of acoustic transients. *Neuroreport* 8, 1761–1765
- 20 Fiez, J.A. et al. (1995) PET studies of auditory and phonological processing - effects of stimulus characteristics and task demands. *J. Cogn. Neurosci.* 7, 357–375
- 21 Temple, E. et al. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: Evidence from functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13907–13912
- 22 Poldrack, R.A. et al. (2001) Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *J. Cogn. Neurosci.* 13, 687–697
- 23 Belin, P. et al. (1998) Lateralization of speech and auditory temporal processing. *J. Cogn. Neurosci.* 10, 536–540
- 24 Schonwiesner, M. et al. (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528
- 25 Zatorre, R.J. and Gandour, J.T. (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 1087–1104
- 26 Jamison, H.L. et al. (2006) Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* 16, 1266–1275
- 27 Zaehle, T. et al. (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *Eur. J. Neurosci.* 20, 2447–2456
- 28 Beauvois, M.W. (1998) The effect of tone duration on auditory stream formation. *Attn. Percept. Psychophys.* 60, 852–861
- 29 Scott, S.K. et al. (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406
- 30 Eisner, F. et al. (2010) Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30, 7179–7186
- 31 Narain, C. et al. (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. cortex* 13, 1362–1368
- 32 McGettigan, C. et al. (2012) An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *J. Cogn. Neurosci.* 24, 636–652
- 33 McGettigan, C. et al. (2012) Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia* 50, 762–776
- 34 Mummery, C.J. et al. (1999) Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457

- 35 Liebenthal, E. *et al.* (2005) Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631
- 36 Agnew, Z.K. *et al.* (2011) Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *J. Cogn. Neurosci.* 23, 4038–4047
- 37 Jacquemot, C. *et al.* (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546
- 38 Friederici, A.D. (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84
- 39 Friederici, A.D. *et al.* (2010) Disentangling syntax and intelligibility in auditory language comprehension. *Hum. Brain Mapp.* 31, 448–457
- 40 Herrmann, B. *et al.* (2012) Dissociable neural imprints of perception and grammar in auditory functional imaging. *Hum. Brain Mapp.* 33, 584–595
- 41 Obleser, J. *et al.* (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289
- 42 Obleser, J. and Kotz, S.A. (2010) Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640
- 43 Belin, P. *et al.* (2000) Voice-selective areas in human auditory cortex. *Nature* 403, 309–312
- 44 Belin, P. and Zatorre, R.J. (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14, 2105–2109
- 45 Belin, P. *et al.* (2002) Human temporal-lobe response to vocal sounds. *Cogn. Brain Res.* 13, 17–26
- 46 von Kriegstein, K. *et al.* (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn. Brain Res.* 17, 48–55
- 47 Kriegstein, K.V. and Giraud, A.L. (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 22, 948–955
- 48 von Kriegstein, K. *et al.* (2005) Interaction of face and voice areas during speaker recognition. *J. Cogn. Neurosci.* 17, 367–376
- 49 Petkov, C.I. *et al.* (2008) A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374
- 50 Johnsrude, I.S. *et al.* (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155–163
- 51 Meyer, M. *et al.* (2002) fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* 17, 73–88
- 52 Meyer, M. *et al.* (2004) Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang.* 89, 277–289
- 53 Kotz, S.A. and Schwartz, M. (2010) Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399
- 54 Rosen, S. *et al.* (2011) Hemispheric asymmetries in speech perception: sense, nonsense and modulations. *PLoS ONE* 6, <http://dx.doi.org/10.1371/journal.pone.0024672>
- 55 Diehl, R.L. (2008) Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 965–978
- 56 Kluender, K.R. and Alexander, J.M. (2010) Perception of speech sounds. In *The Senses: a Comprehensive Reference* (Basbaum, A.I. *et al.*, eds), pp. 829–860, Elsevier
- 57 Ohala, J.J. (1983) The origin of sound patterns in vocal tract constraints. In *The Production of Speech* (MacNeilage, P.F., ed.), pp. 189–216, Springer-Verlag
- 58 Stilp, C.E. *et al.* (2010) Auditory color constancy: calibration to reliable spectral properties across nonspeech context and targets. *Attn. Percept. Psychophys.* 72, 470–480
- 59 Ulanovsky, N. *et al.* (2003) Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398
- 60 Gagnepain, P. *et al.* (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2012.1002.1015>
- 61 Schwartz, M. *et al.* (2012) Temporal aspects of prediction in audition: Cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.* 83, 200–207
- 62 Shankweiler, D. and Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. *Q. J. Exp. Psychol.* 19, 59–63
- 63 Studdert-Kennedy, M. and Shankweiler, D. (1970) Hemispheric specialization for speech perception. *J. Acoust. Soc. Am.* 48, 579–594
- 64 Cutting, J.E. (1974) Left-hemisphere mechanisms in speech-perception. *Percept. Psychophys.* 16, 601–612
- 65 Schwartz, J. and Tallal, P. (1980) Rate of acoustic change may underlie hemispheric-specialization for speech-perception. *Science* 207, 1380–1381
- 66 Giraud, A.-L. *et al.* (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134
- 67 Luo, H. and Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010
- 68 Ahissar, E. *et al.* (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372
- 69 Luo, H. *et al.* (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8, e1000445
- 70 Kerlin, J.R. *et al.* (2010) Attentional gain control of ongoing cortical speech representations in a 'cocktail party'. *J. Neurosci.* 30, 620–628
- 71 Obleser, J. and Weisz, N. (2011) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb. cortex* <http://dx.doi.org/10.1093/cercor/bhr325>
- 72 Palva, S. and Palva, J.M. (2007) New vistas for alpha-frequency band oscillations. *Trends Neurosci.* 30, 150–158
- 73 Jensen, O. and Colgin, L.L. (2007) Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* 11, 267–269
- 74 Palva, J.M. *et al.* (2005) Phase synchrony among neuronal oscillations in the human cortex. *J. Neurosci.* 25, 3962–3972
- 75 Hanslmayr, S. *et al.* (2011) The relationship between brain oscillations and BOLD signal during memory formation: a combined EEG-fMRI study. *J. Neurosci.* 31, 15674–15680
- 76 Josse, G. *et al.* (2009) Predicting language lateralization from gray matter. *J. Neurosci.* 29, 13516–13523
- 77 Poldrack, R.A. (2007) Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.* 2, 67–70
- 78 Misaki, M. *et al.* (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118