Do temporal processes underlie left hemisphere dominance in speech perception?

Sophie K Scott[*] and **Carolyn McGettigan**
Institute for Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK

Abstract

It is not unusual to find it stated as a fact that the left hemisphere is specialized for the processing of rapid, or temporal aspects of sound, and that the dominance of the left hemisphere in the perception of speech can be a consequence of this specialisation. In this review we explore the history of this claim and assess the weight of this assumption. We will demonstrate that instead of a supposed sensitivity of the left temporal lobe for the acoustic properties of speech, it is the right temporal lobe which shows a marked preference for certain properties of sounds, for example longer durations, or variations in pitch. We finish by outlining some alternative factors that contribute to the left lateralization of speech perception.

Introduction

The localization of cortical fields important in speech perception and production represented the first cognitive specializations to be discovered in the human brain: Broca identified a role for the posterior third of the left inferior frontal gyrus in speech production, and Wernicke delineated a role for the left superior temporal gyrus in speech perception (Broca, 1861; Wernicke, 1874). Functional imaging techniques and developments in our understanding of primate auditory anatomy have enabled us to integrate the functional and anatomical aspects of Broca's and Wernicke's area into more general models of the speech perception system (Hickok & Poeppel, 2000, 2004; G. Hickok & D. Poeppel, 2007; Rauschecker & Scott, 2009; S. K. Scott & Johnsrude, 2003). There is continued debate, however, as to why there is a clear dominance (in the human brain) of the left hemisphere in speech perception and production. In terms of speech perception, there are two general approaches to the issue of speech perception lateralization (Robert J. Zatorre & Gandour, 2008). One popular approach (Poeppel, 2003; R. J. Zatorre & Belin, 2001) posits domain general principles that underlie the hemispheric asymmetry, namely that acoustic biases between the two hemispheres lead to preferential processing of aspects of speech on the left (though Poeppel's account expressly considers prelexical processing of speech to be a bilateral phenomenon, in which the left temporal lobe is specialized for temporal processing which is considered to favour the phoneme). By Zatorre's example, therefore, the left temporal lobe is not specialized for speech because of its linguistic status but rather because of aspects of its acoustic structure. Other, more domain specific accounts, associate left

[*]corresponding author: sophie.scott@ucl.ac.uk, tel: +44 (0) 20 7679 1144.

hemispheric dominance in speech perception with linguistic mechanisms – such as syntactic processing (Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010)

We have recently published a review (Carolyn McGettigan & Sophie K. Scott, 2012) critically investigating the idea that acoustic sensitivities lead to the preferential processing of certain sounds (and therefore aspects of speech) in the left temporal lobe (Poeppel, 2003; R. J. Zatorre & Belin, 2001). The aim of the current paper is to take a more historical approach to the problem by evaluating the essential claim that the left hemisphere is specialized for rapid, or temporal processing. We will demonstrate not only that this evidence is often a misreading of the original papers, or that the papers overstate their claims, but that many of the same errors of interpretation have dogged the functional imaging literature that has grown up around this topic. In this paper we will use speech perception to refer to perceptual processing of heard speech, and speech comprehension to refer to wider, amodal linguistic systems. Implicit in this framework is the assumption that perceptual processes form part of a larger comprehension network.

Time in temporal order judgments

In 1963, Efron published a highly influential paper addressing temporal processing in aphasic patients (Efron, 1963). By 'temporal processing', Efron meant explicit judgments of temporal order: he argued that as the elements of speech need to be in the correct order for successful perception, temporal ordering might be central to the accurate decoding of speech, and that disorders of speech perception might involve a breakdown in such accurate sequencing. To investigate this, he worked with a group of patients who had problems with speech perception and production. He found that the patients made temporal order judgments errors over a wider range of stimulus onset intervals than the control subjects – the threshold interstimulus intervals yielding 75% correct performance with auditory sequences were 70ms for controls, 400ms for those with expressive aphasia, and 140ms for those with receptive aphasia. The threshold interstimulus intervals yielding 75% correct performance with visual sequences were 80ms for controls, 100ms for expressive aphasic patients, and 160ms for the receptive aphasic patients. In other words, to report the stimuli in the correct order, both groups of aphasic patients needed longer interstimulus intervals than the control subjects, for both auditory and visual sequences. This paper has been very influential, showing that non-linguistic tests can reveal interesting processing problems in patients with ostensibly solely linguistic problems.

There were some oddities in the data: for example, the 'expressive aphasics' (i.e. those with a speech production problem) showed much longer thresholds for the auditory sequences than the receptive aphasics (i.e. those with a speech perception problem), and all the patients had problems with intervals that are relatively long (e.g. 100ms and upwards) when compared to the kinds of short temporal windows that people later would often consider to be both essential to speech and preferentially processed on the left (e.g. 25ms; Poeppel, 2003). However this later point was not a problem for Efron's thesis, as he was addressing the ordering of the stimuli, not the precise time scales involved. Efron considered that the kinds of temporal information he was tapping into would permit the sounds of speech to be heard in the correct order. However there are limitations to this study: as no right

hemisphere damaged group was tested, it is hard to know the extent to which these findings demonstrate left lateralized processes that are specific to temporal order judgments, or more generally, whether these results are dominated by perceptual or response selection deficits.

Tallal and Piercy also used temporal order judgments as a way of investigating mechanisms which might underlie speech processing deficits (Tallal & Piercy, 1973). They were investigating children with a developmental aphasia, who would probably have a diagnosis of specific language impairment (SLI) nowadays. They reported that relative to a control group, children with speech problems needed longer intervals between the auditory stimuli before they could accurately report which came first. Tallal and Piercy made more specific claims than Efron, directly linking the children's speech problems to a deficit in *rapid* temporal processing, rather than in sequencing of the information. Again it is striking that the children with language problems diverged from the control group at relatively long inter-onset intervals, on the order of 300ms. This does make the invocation of 'rapidly presented sequences' more problematic, since 300ms is longer than the average syllable duration (round 200ms), and much longer than the duration of a phoneme.

Time and dichotic listening

In the absence of anatomical and functional imaging techniques, it was hard to determine the extent to which any perceptual, motor or cognitive processes might be lateralized in the intact human brain. Efron's participants had 'unequivocal' left hemisphere lesions (based on a neurological investigation, we assume), but assumptions about the lateralization of the processes investigated by Tallal and Piercy involves a more indirect inference about left hemispheric mechanisms, due to the children having language problems. However, researchers could determine hemispheric specialization or preferences by using dichotic listening tasks, which take advantage of fact that there are projections from each to the contralateral hemisphere, such that the left ear projects to the right side of the brain and vice versa (although the incomplete decussation of the human ascending auditory pathways means that there are also ipsilateral connections). In dichotic listening tasks, superior performance of a stimulus presented to one ear (vs the other ear) is taken as an indication of preferential processing of that stimulus or class of stimuli in the contralateral hemisphere (Kimura, 1961).

Dichotic listening appeared to be an excellent tool for addressing the hemispheric asymmetry of speech and sound process in intact human brains. Subsequent work has indicated that dichotic listening can be a good way of looking at large scale differences between the left and right temporal lobes in terms of auditory processing for example, dichotic listening performance has been shown (in combination with fMRI studies) to be a better predictor of language lateralization than handedness (Van der Haegen, Westerhausen, Hugdahl, & Brysbaert, 2013). However, dichotic listening can be very sensitive to precise task constraints (partly because the ipsilateral connections are inhibitory, so what is played in the other ear can have an effect on the profile of asymmetries seen) and so it has not always been as reliable as other techniques. For example, the presence (or absence) of noise at the ear which is not having target stimuli presented to it can have a marked effect on the

extent of an apparent ear advantage (Studdert-Kennedy & Shankweiler, 1981). This can lead to apparently noisy and/or variable ear advantage studies.

With this proviso in mind, several studies have shown that a right ear advantage (REA) can be shown to speech sounds in dichotic listening paradigms. For example, Shankweiler and Studdert-Kennedy (1967) showed that the identification of synthetic consonant-vowel combinations was better for stimuli presented to the right ear than to the left. In contrast, there was no ear advantage for synthetic vowel sounds with steady-state spectra.

This finding was extended to comparable speech and non speech stimuli by James Cutting (Cutting, 1974), who contrasted ear advantages for synthetic consonant-vowel, synthetic vowel, and sine wave equivalents. The results were discussed in terms of a right ear advantage for stimuli with a rapid spectral change (e.g. the consonant-vowel and matched sine wave "consonant-vowel" stimuli), relative to the two spectrally unchanging stimuli (vowels and matched sine wave "vowel" stimuli). However, although there is an overall REA reported for the stimuli with rapid spectral change (based on a t-test comparison, collapsed across speech and non-speech stimuli trials), a closer inspection of the data (replotted here in Figure 1) shows that there was no REA for the sine wave "consonant-vowel" stimuli. Moreover, neither of the non-speech sine tone stimuli were significantly lateralized, in terms of right or left ear advantage. To satisfy Cutting's interpretation, the REA should have been present for all stimuli with transitions, both speech and non-speech sine tone stimuli alike. Instead, these results suggest an interaction between the presence of the transitions and the linguistic status, where there was only an REA for the speech-like consonant-vowel stimuli.

A further dichotic listening study Schwartz and Tallal (1980) reported that an REA could be shown for an identification task using synthesised consonant-vowel syllables (with plosive onsets). This numerical (but not statistically established) REA was reduced, though still apparent, for the plosive stimuli where the formant transitions had been extended from 40ms to 80ms. This was taken to show that the left hemisphere was optimized for the processing of stimuli with very short formant transitions. This was an important development, as it associated the left hemispheric preference for speech to non-verbal aspects of the nature of the stimuli being processed, rather than to overall faster processing speed in the left hemisphere, as suggested by Cutting (1974). Schwartz and Tallal interpreted this effect as showing a left temporal lobe *advantage* for rapid transitions: this interpretation was criticized by Studdert-Kennedy and Shankweiler (1981) who provided a robust retort in which they pointed out that most speech sounds (fricatives, nasals, affricates, liquids) involve spectral transitions over much longer timescales than 40ms, and are no less 'critical' to the perception of speech than the faster spectral changes in plosives. They also added that a difference in the degree of an REA does not necessarily mean differences in the degree of left hemisphere engagement. We would only add that as non-speech stimuli were not presented, it is not clear that this finding can be assumed to be independent of a speech modality. Furthermore, the characterization of 80ms as 'long' in temporal terms is in direct contradiction with the range of temporal windows over which the groups who showed problems with temporal order judgments (TOJs) for non-speech stimuli (Efron, 1963; Tallal & Piercy, 1973).

In a brace of critiques of the rapid temporal processing approach to SLI, Studdert-Kennedy and Shankweiler (1981) and Studdert-Kennedy and Mody (1995) point out several of the shortcomings of this approach, not least the shifting senses of the concepts of, *temporal* and *temporal processing*, which are sometimes used in a mutually exclusive fashion (e.g. to refer to brief transitions within a stimulus, or to the perceived order of separate stimuli). This can be seen in the more recent studies, where temporal is a phrase used to described aspects of the sampling rate, or window size, of perceptual processes (Poeppel, 2003), or to connote aspects of time varying properties of the amplitude envelope of a sound (Obleser et al, 2008), or speed with which auditory events occur (Zatorre and Belin, 2001. Studdert-Kennedy and Mody also point out that it is taken as axiomatic that rapid temporal processing is central to speech perception, without any clear evidence to actually support this assertion. Indeed, by Studdert-Kennedy and Shankweiller's estimation (Studdert-Kennedy & Shankweiler, 1981), in spoken English rapid formant transitions account for perhaps 20% of the speech signal. It's fair to say that these criticisms have largely been ignored, such that subsequent cognitive neuroscience investigations, using different functional imaging tools, have still largely investigated speech using place of articulation in plosives (as mentioned in. D'Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011; S. K. Scott & Wise, 2004) Along these lines, we have described elsewhere some *a priori* reasons why rapid temporal processing might not be a good way of characterizing the speech signal (McGettigan, Evans, et al., 2012; Carolyn McGettigan & Sophie K. Scott, 2012), including the fact that many speech sounds are longer than the plosive transitions described (e.g. nasals, vowels, liquids, fricatives, affricates), and that co-articulation effects mean that information in speech is spread over adjacent speech sounds, such that information in speech is not located solely at the level of the individual phonemes (e.g. Kelly & Local, 1989). Certainly we would argue that, for any account of speech perception, either domain general or domain specific, aspects of spectral processing need to form a central aspect, given the dependence of speech on this kind of modulation. For example, to perceive the difference between /ba/ and /da/ requires spectral processing, as the temporal characteristics are identical (Studdert-Kennedy & Mody, 1995).

Evidence from functional imaging

Despite the empirical and theoretical shortcomings of the temporal processing account of speech and language processing, in both normal and disordered populations, this approach has enjoyed great influence in the rapidly developing field of cognitive neuroscience, and several early papers on the neural basis of speech perception uncritically cited this approach as underlying left temporal lobe responses to speech (e.g. Mummery, Ashburner, Scott, & Wise, 1999; Wise et al., 1991). Problems started to appear, however, when researchers started to actually test the theory that there would be a left hemisphere preference for rapid auditory processing, using functional imaging. Unlike patient testing and dichotic listening experiments, functional imaging offers the opportunity to directly interrogate both the presence (or absence) of hemispheric asymmetries, and a chance to delineate their anatomical location. Fiez and colleagues (Fiez et al., 1995) reported a left frontal operculum sensitivity to temporal aspects of speech and tone processing. Johnsrude and others (Johnsrude, Zatorre, Milner, & Evans, 1997) showed a left hemisphere dominance for the processing of pure tones with shorter (30ms) or longer (100ms) changes in pitch. A direct

contrast of the two pitch transition conditions showed greater activation for the shorter stimuli in left orbitofrontal cortex, left fusiform gyrus, and right cerebellum, and a greater response in the right dorsolateral frontal cortex for the longer transitions. Both of these studies show some evidence for a left hemisphere dominance to temporal properties of sound, or stimuli with shorter transitions, but the activation is not associated with sensory aspects of auditory processing, that is, the peaks reported all fall outwith auditory cortical fields.

A study from Belin and colleagues (Belin, Zilbovicius, Crozier, Thivard, & Fontaine, 1998) used synthesized speech-like stimuli with fast (40ms) or slow (200ms) onset transients, which did not form meaningful stimuli for the listeners (i.e. they sounded like non-native speech). The steady state portion of the stimulus did not vary in duration, so the faster stimuli were also shorter. In contrast to the previous findings (Fiez et al., 1995; Johnsrude et al., 1997), this study did reveal stimulus-specific effects in the temporal lobes, possibly due to the use of more spectrally complex 'speech-like' stimuli, or to the wider range of transitions/durations used. A direct contrast of the fast and slow stimuli showed no specific left hemisphere preference for the faster stimuli: instead, the left temporal lobe responded equally to the fast-transitions and the slow-transitions stimuli, while the right temporal lobe was selectively activated by the longer slower-transition stimuli. This study has been widely interpreted as showing a selective response of the left auditory areas for faster stimuli, while in fact it is the right temporal lobe which is showing a selective response for longer stimuli with slower onset transients. From a phonetic point of view, one might have predicted that if the left temporal lobe is more activated by speech, then it should respond to both fast and slow transients, as speech contains informative modulations across a range of time scales, fast and slow (Rosen, Wise, Chadha, Conway, & Scott, 2011; Studdert-Kennedy & Shankweiler, 1981). The right temporal lobe selectivity was, however, a novel and important early demonstration that right auditory fields do show some acoustic sensitivity to certain properties of sounds (here, longer durations and/or slower transitions).

Theories of auditory sensitivities underlying asymmetries in speech perception

Around the start of the new millennium, a couple of different but related accounts (Poeppel, 2003; R. J. Zatorre, Belin, & Penhune, 2002) were advanced about the cortical processing of human speech, building on the concepts of a left hemisphere dominance for rapid and/or temporal processing which as described above had been developed in the clinical (Efron, 1963; Tallal & Piercy, 1973, 1974; Tallal & Stark, 1981) and speech perception literature (Cutting, 1974; Schwartz & Tallal, 1980). Both of these theories provided mechanistic, neuro-biologically instantiated accounts of a left hemisphere dominance for specific, non-linguistic acoustic properties of sounds, and both did so by proposing complementary processing differences between left and right auditory fields. Neither of these theories was making claims about how these asymmetric auditory sensitivities would affect post-perceptual linguistic processes such as syntax.

Robert Zatorre was expressly accounting for a left hemispheric dominance in the perceptual processing of speech, and a right hemispheric dominance in the perceptual processing of music (R. J. Zatorre et al., 2002). He postulated that, as in the construction of spectrograms,

left auditory fields code fine temporal resolution of acoustic stimuli, with correspondingly poorer spectral detail. Zatorre's model predicts a contrasting profile in right auditory areas, which code spectral detail in fine resolution but with correspondingly poorer temporal resolution (R. J. Zatorre et al., 2002). Zatorre's account makes the fundamental error of assuming that the human auditory system preserves detailed spectral information in the ascending auditory pathway and into cortical fields: instead, human auditory processing involves the loss of a lot of spectral detail at the level of the cochlea, and this information cannot be retrieved at the cortex. The loss of spectral detail is a result of the human auditory system representing lower frequency information in more detail than higher frequency information: this is why the kinds of spectral compression used to make mp3 versions of sound files has minimal perceptual effects on the acoustic quality of the sounds, as this process involves removing information from higher frequencies, at which the spectral resolution at the cochlea is poorer. It is also the case that speech perception not only does not require fine temporal resolution, but it does require some good spectral resolution; the processing of lower frequency spectral information, as it evolves over time, is critical in speech perception (Rosen et al., 2011; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

The Asymmetric Sampling in Time model (Poeppel, 2003) posits that prelexical speech perception happens bilaterally in the brain, but that the left and right temporal lobes are specialized to deal with different kinds of acoustic information. This is a predominantly temporal account in which the left and right auditory areas differ in the size of the windows over which they preferentially sample information. Poeppel identifies temporal modulations in the gamma (~40Hz) and theta (and alpha; ~4-10Hz) ranges as critical for the processing of acoustic changes in speech within the range of the phoneme and the syllable, respectively. Instantiated within the brain, the theory proposes that cells in the auditory cortices of both hemispheres exhibit specific tuning windows for gamma- and theta-range information (Poeppel, 2003). However, the asymmetry emerges because the left auditory cortex is hypothesized to contain a predominance of gamma-tuned populations, and the right auditory cortex a complementary predominance of theta-tuned cells (though the precise specifications of the hypothesized temporal windows have themselves varied across (and within) subsequent papers (Obleser, Herrmann, & Henry, 2012). It's worth stressing that, *temporal* in Poeppel's approach, is making a specific prediction about the time scale over which information is expressed in speech production and integrated in speech perception. Further, these different time scales of information (short vs long) are accessed by cortical processes which integrate the information over smaller or larger time windows. These processes in turn are indexed by different cortical oscillations. It is fair to note that the perceptual events that are encoded within these windows need not themselves be purely 'temporal', but that they contain structure at this temporal grain. In Zatorre's theory, *temporal* refers to a property of sounds which is sometimes instantiated in the theory as brevity, or rapidity, in transitions and sometimes as associated with properties of the amplitude envelopes of sounds.

Testing the theories of hemispheric asymmetries - Where is the left hemispheric preference for temporal processing?

As in the earlier paper by Schwartz and Tallal (1980), the approaches of both Zatorre (R. J. Zatorre et al., 2002) and Poeppel (2003) take as a basic, uncontroversial assumption, that rapid auditory information, on the time scale of the formant transitions for plosives, is central to speech perception. They also typically de-emphasied the role of spectral processing (Giraud & Poeppel, 2012; R. J. Zatorre et al., 2002). Both of these arguments are contentious (C. McGettigan & S. K. Scott, 2012). However, it could be argued that if the left hemisphere dominance in speech perception is based on a non-linguistic acoustic preference or selectivity, that non-linguistic property of the sound would need not be related in any transparent or obvious way to speech – for example, it could represent an important precursor to speech related processing. Although both theories explicitly relate speech perception to rapid auditory processing, this link might not be necessary if a genuine preference can empirically be demonstrated – that is a sensitivity in the response of left auditory cortex to acoustic properties could still be important in the absence of a direct link to aspects of the acoustic properties of speech.

However, demonstrations of preferential auditory responses in the left temporal lobe have remained somewhat elusive. On one hand, Zaehle and colleagues (Zaehle, Wustenberg, Meyer, & Jancke, 2004) reported an unambiguous left auditory cortex response to a gap detection task, using the perception of very brief (8/32ms) gaps in noise. Although this activation was medial and dorsal to the typical location of neural responses to linguistic aspects of speech (e.g. McGettigan, Evans, et al., 2012), the study clearly reveals a left hemispheric dominance in gap detection, a task which is genuinely temporal in nature. Such clarity has not been possible with many other studies, due to partly to complexities in the definition of the tasks/stimuli, and also due to the ways that the data are analysed and interpreted.

Zatorre and Belin (R. J. Zatorre & Belin, 2001) contrasted 'temporal' and 'spectral' aspects of stimuli. The temporal manipulations were made by varying the average ISI between two sine tones at two different frequencies to create a continuum of interstimulus intervals, across one fixed pitch interval. The spectral manipulations were made by varying the number of intervening pitches to create a continuum of the distribution of spectral elements (from many to fewer). This 'spectral' continuum is something of a misnomer; as all the tones are pure sine tones, the instantaneous spectrum of the sequences is highly similar in all conditions. This condition is better conceived as one where the number of different elements varies across the sequences, from few pitch intervals to many (which is also consistent with the way in which 'spectral' is often used synonymously with 'pitch' in the paper). The results of this study showed bilateral responses to both 'temporal' and 'spectral' change: the temporal effects were in or near primary auditory cortex, and the 'spectral' effects were in anterior temporal lobe fields. One of the right STS peaks showed a correlation only to the 'spectral' changes: this is congruent with evidence showing that right temporal fields are important in the perception of pitch changes and melody (Johnsrude, Penhune, & Zatorre, 2000). In contrast, there was no significant difference between the response of the left and the right Heschl's gyrus to the temporal variation. As in the Belin et al. study (1998),

therefore, a selective response can be seen in the right temporal lobe to an aspect of the stimuli (here, increasing amount of pitch detail in a sequence), but no selectivity (in either hemisphere) to the temporal manipulations. A second study by Jamison and colleagues (Jamison, Watkins, Bishop, & Matthews, 2006) using the same stimuli and an fMRI paradigm replicated the right anterior STS response to the pitch interval detail manipulation, and the bilateral response to the temporal manipulation. A direct comparison of the size of the effect in the right and left STG to the temporal manipulation found a significant difference, with a greater response on the left than the right. However, the response to the temporal manipulation was significant in both left and right STG. A similar profile has been reported in another paper: Schoenwiesner and colleagues (Schonwiesner, Rubsamen, & von Cramon, 2005) reported that increasing the spectral complexity of sounds led to a greater response in the right anterolateral parabelt auditory field, while increasing the rate of change of temporal information led to neural responses which did differ significantly between the left and right temporal lobe.

Another study, directly testing the AST hypothesis, varied the duration of elements in non-speech sequences to create a spectrum of sequences containing different element duration (from longer to shorter): each element also either contained some spectral/pitch change or not. (Boemio, Fromm, Braun, & Poeppel, 2005) Overall, the right STS was highly sensitive to the duration of elements: as they become longer, the right STS was more active. The predicted corresponding response – left STG/STS showing a greater response to shorter duration stimuli – was not found. Instead, the left temporal lobe simply showed a weaker but still increased response to increases in element duration. The lack of the predicted left auditory response was not discussed in the paper. A similar effect has also been demonstrated in an fMRI study of the cortical responses to increasing temporal windows for spectral coherence in sounds (Overath, Kumar, von Kriegstein, & Griffiths, 2008): in this paper, the differences between the lack of a left temporal lobe sensitivity to shorter temporal windows over which structure evolved in sounds and the predictions of the Zatorre and Poeppel models was discussed. This is often not the case, and the datasets clearly showing acoustic preferences in the right hemisphere are often interpreted as if a corresponding, complementary left hemisphere preference must be present, since the neural response profiles are different (Belin et al., 1998; Boemio et al., 2005; Schonwiesner et al., 2005; R. J. Zatorre & Belin, 2001).

With the exception of Overath and colleagues (Overath et al., 2008), it seems strange that this marked *lack* of a left hemisphere preference for temporal aspects of sound should be so rarely commented on, or indeed, that such a preference could be still somehow perceived as present (e.g. Boemio et al., 2005; R. J. Zatorre & Belin, 2001). There has been a general tendency to report left hemisphere findings in a way that aligns with the dominant theories but which may not fully reflect the overall pattern of results. For example, an fMRI study in which speech stimuli were either noise vocoded to different numbers of channels (a spectral-temporal manipulation which predominantly affects spectral detail) or the amplitude envelope was smoothed within each channel to different amounts (a spectral-temporal manipulation in which more of less amplitude envelope modulation detail, or 'temporal' detail, is preserved). The results revealed that the spectral modulations had the biggest

effects on the speech intelligibility and drove activation in the left STG/STS more than the temporal manipulation, which had less impact on intelligibility. Nonetheless the results were described as showing a left hemisphere selectivity for temporal processing, as the response to the temporal manipulation was greater on the left than the right (Obleser, Eisner, & Kotz, 2008). The response to the 'spectral' manipulation was greater on the left that it was to the temporal manipulation, so another plausible interpretation was that the left temporal lobe was more sensitive to spectral manipulations than to temporal manipulations; or that the left temporal lobe was most sensitive to manipulations which affected the intelligibility of the speech sequences.

There has been some development of these positions. In a 2008 paper, Zatorre and Gandour argued that we should move beyond a domain-general vs domain specific dichotomy in how we approach the left hemisphere dominance for speech processing (Robert J. Zatorre & Gandour, 2008): however, by this they still hold that a non speech specialization for rapid acoustic processing underlies the perceptual processing of speech, and that domain specific mechanisms are involved later in the processing of spoken language. This is startlingly at odds with the empirical evidence, given the minority of papers that empirically demonstrate a left temporal lobe preference or selectivity for temporal aspects of sound.

It has also been suggested that the auditory processing asymmetries might be mild or weak, leading to tendencies, rather than overall clear differences in processing (Robert J. Zatorre & Gandour, 2008). This is similarly difficult to square with the evidence: the right temporal lobe preference for sounds with pitch variation/longer durations is robust, easily demonstrated and unambiguous (Belin et al., 1998; Boemio et al., 2005; Schonwiesner et al., 2005; R. J. Zatorre & Belin, 2001), and the left temporal lobe dominance in speech perception is also easily shown at the level of phonotactic structure (Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003), phonemes (Agnew, McGettigan, & Scott, 2011), consonant-vowel combinations, words and sentences (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; McGettigan, Evans, et al., 2012; McGettigan, Faulkner, et al., 2012; Mummery et al., 1999; Narain et al., 2003; S. K. Scott, Blank, Rosen, & Wise, 2000). Why would an auditory preference on the left, but not the right, be weak? What use would a weak preference serve, if the temporal cues posited are simultaneously so critical to speech?

Current models: the role of neural oscillations

David Poeppel's position has modified over the decade since his AST model was first proposed (Poeppel, 2003). Although he still strongly emphasizes temporal information in speech and preferential neural processing of sound within two general time windows (the specific details of which are permitted some flexibility; Ghitza, Giraud, & Poeppel, 2013; Obleser et al., 2012), recent papers have tended to focus less on, *where* and rather more on *how* and, *when* speech is processed in auditory cortex. The role of hemispheric asymmetries has not been lost but may be somewhat less emphasized in this more recent work. Central to the current model (Giraud & Poeppel, 2012) is the relationship between neural oscillations in the theta and gamma bands, and their parsing of, and entrainment to, the auditory speech signal. Specifically, the model holds that theta phase-resetting and coupling to the amplitude

envelope of speech, around the rate of syllable production, in turn engages a nested modulation of gamma power, which subserves the segmentation and perception of finer-grained information on the timescale of phoneme. Theta is seen as the "master oscillator" (Ghitza et al., 2013), a view which is based on a preponderance of neural responses to sound within this window in the human brain, as well as in the acoustic signal of speech (in the fluctuations of the amplitude envelope at the rate of syllable production). Peelle & Davis (2012) outline experimental evidence that suggests the importance of envelope cues to the comprehension of speech, although some authors have cautioned an over-emphasis on the amplitude envelope that would overlook other important spectro-temporal cues (Cummins, 2012; Obleser et al., 2012; S. Scott & McGettigan, 2012).

There are some outstanding questions about the interplay between acoustics and neural responses, and the degree to which these must be matched in order for speech to be perceived as intelligible. Initial studies investigating a theta phase-locked response to the amplitude envelope of speech showed that compromising the fidelity of envelope following (achieved experimentally by degrading the acoustic signal) led to impaired speech comprehension (Ahissar et al., 2001; Luo & Poeppel, 2007). The view that decreased phase-locking to the envelope was causal of impaired perception was clearly stated within the recent description of the model by Giraud and Poeppel (2012). However, the same authors have also argued that the relationship between theta and the envelope can be purely acoustic – Howard and Poeppel (2010) showed that theta phase-locked responses can discriminate between individual sentences when both naturally presented and when time-reversed (and therefore unintelligible) - and a recent perspective (Ghitza et al., 2013) notes that envelope cues alone are sufficient for a strong theta response, but that they are enhanced by the presence of additional acoustic-phonetic information in the signal (see also J. E. Peelle, Gross, & Davis, 2012).

We see several key issues that must be addressed in future research testing the Giraud & Poeppel (2012) model:

1. Can we be sure that theta is the "master" response for tracking the speech envelope? Nourski and colleagues (2009) found envelope-following responses in high gamma power (70-250 Hz) for time-compressed stimuli with compression rates high enough to render the stimuli unintelligible, while low-frequency evoked responses lost fidelity to the envelope at these higher compression rates. Could it be that the limitations of M/EEG to measure responses in the higher frequency ranges (cf the electrocorticography (ECoG) method used by Nourski et al., 2009) are distorting our prevailing view toward low-frequency responses? Related to this, there is the issue that a relatively sluggish theta response, taking several hundred milliseconds to entrain (Abrams, Nicol, Zecker, & Kraus, 2008; Ghitza, 2012, 2013), would presumably struggle to adequately encode the onsets of speech stimuli adequately in the perceptual system. "for single isolated In a recent commentary, that based models do Ghitza not provi (2013) suggests de any addi words, oscillatortional insights into our understanding of how sub-word units are decoded. This is so because the duration of the stimulus is too short to allow entrainment." (p. 4).

2. To what extent must any neural account of speech perception prioritize the matching of neural and acoustic oscillations? While the predominance of theta-range information in the speech signal, and in the brain's response to it, seems an enticing prospect for a mechanistic account, this breaks down if we consider the evidence from Nourski and colleagues (2009) for high-frequency envelope tracking in the 70-250 Hz range, and Peña and Melloni's (2012) main finding of preferential gamma enhancement in the 55-75Hz range for intelligible sentences from the participants' native language – here there is no neat alignment of the neural oscillations with syllabic or phonemic cues on the order of 150-250ms or 20-40ms, respectively (Poeppel, 2003). If, as proposed in the recent perspective offered by Ghitza, Giraud & Poeppel (2013), the precise frequency range of these faster responses is no more than a ratio factor of the "master", then any evidence for high-frequency tracking of the speech signal should still be supported by equivalent fidelity in the low-frequency tracking responses driving the model (see point 1 above).

3. Finally, if theta responses are locked to the stimulus but insufficient for intelligibility (Nourski et al., 2009), what is the mechanism for observed enhancements in theta related to increased comprehension, and what other responses are permitted to contribute to this model? Work on the neural correlates of degraded speech comprehension from Obleser and Weisz (2012) identified an increasing alpha power suppression, in inverse proportion with theta power enhancement, in response to speech of increasing intelligibility (with roughly consistent envelope cues). Peña and Melloni (2012) relate mid-gamma responses to intelligible speech to a role for gamma in the coordination of neural responses for the comprehension of multi-word linguistic messages. In response to a critique of the Giraud & Poeppel (2012) model, Ghitza, Giraud and Poeppel (2013) suggest that alpha-theta or delta-theta interactions may support the use of acoustic-phonetic cues to comprehension. In the context of the nested theta-beta-gamma model put forward by Ghitza (2011) and described by Ghitza, Giraud & Poeppel (2013), the inclusion of delta and alpha interactions would essentially mean that little information between 1-160Hz is *excluded* from contributing to the model, making it rather less discrete than previously suggested. This takes us back to the point made in the previous section – is there a fundamental necessity for an alignment of acoustic cues and neural oscillations, or could these models be just as well served without discrete acoustic dependencies? If theta-gamma coupling is a central property of human cerebral cortical function, then is it necessary to yoke processes in these frequency ranges to acoustic events (Lisman & Jensen, 2013)?

At the time of writing, the discussion of neuronal oscillations in speech perception/comprehension is active and ongoing, and it is an exciting period of openness for the field. It remains to be seen the extent to which any emergent models can be instantiated in terms of gross neuroanatomy – some results indicate differences in the nature of the oscillatory responses across hemispheres (though not always in the direction predicted by AST; Millman, Woods, & Quinlan, 2011) and along the auditory processing hierarchy within-hemisphere (primary versus association auditory cortices; Morillon, Liegeois-Chauvel,

Arnal, Benar, & Giraud, 2012; J. E. Peelle et al., 2012). We suggest that the further development of the field will see the greatest benefits from collaborative approaches that take account of all sides of the question – neurophysiological, computational, psychoacoustic, phonetic, and linguistic.

Is speech left lateralized in terms of perception?—Some theorists in speech and language processing would suggest that the 'elephant in the room' when discussing hemispheric asymmetries in speech perception is the perspective that many of the mechanisms of speech perception are not a lateralized phenomena, and that there is therefore perhaps no need to postulate meaningful hemispheric acoustic asymmetries.. This bilaterality argument has been made predominantly by Hickok and Poeppel (Hickok & Poeppel, 2000; Gregory Hickok & David Poeppel, 2007; though see also Jonathan E. Peelle, 2012). Hickok and Poeppel's position is that prelexical processing of speech is associated with bilateral processing in the dorsolateral temporal lobes: lexical and post lexical processing is a left hemisphere phenomenon. This view is often specified in articles which simultaneously posit Poeppel's (2003) AST model as accounting for the early perceptual (and presumably therefore prelexical) processing of speech, with information on the scale of phonemes being processed on the left and syllable level information being processes on the right (Giraud et al., 2007; Gregory Hickok & David Poeppel, 2007). In other words, the implicit assumption from this perspective is that prelexical processing of speech is bilaterally mediated, but there are still hemispheric asymmetries underlying this – that is, prelexical processing involves different processes in the left and right dorsolateral temporal lobes.

The evidence in favour of the idea that speech perception is mediated bilaterally is the rare disorder 'pure word deafness', in which people can hear sounds, but cannot hear useful structure within the sounds, such that complex information such as phonetic information is unavailable to them. Patients with pure word deafness cannot therefore understand speech (and can struggle with other complex sounds, like music). Pure word deafness can occur after left temporal lobe lesions, but can also be seen after bilateral temporal lobe damage. A patient with a left STG lesion, and a Wernicke's aphasia profile, for example, only presented with pure word deafness following a right STG lesion (Praamstra, Hagoort, Maassen, & Crul, 1991). This is not strong evidence for a truly bilateral, equipotential involvement of both temporal lobes in speech perception, however: pure word deafness is associated with both left hemisphere and bilateral damage, but hardly ever with damage constrained to the right temporal lobe (Griffiths, 2002). Furthermore, the evidence for the kinds of lesions which affect the perception of spoken language has been strongly associated with profiles of left hemisphere damage – it may well be true that the problems associated with such need not be limited to speech perception (Saygin, Dick, Wilson, Dronkers, & Bates, 2003) – but these problems have not been consistently linked to damage which has solely affected the right superior temporal lobe. Specific patterns of deficits can be identified following damage to the right superior temporal lobe, such as voice processing, prosody processing, pitch direction processing (Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010; Johnsrude et al., 2000; Ross, 1981): these problems are typically reported in the context of preserved speech perception, however.

There is also evidence that the right temporal lobe can itself underpin changes following aphasic stroke following damage to the left temporal lobe: Crinion and colleagues (Crinion & Price, 2005) showed that amount of right temporal lobe activation correlated with speech perception performance following a left temporal lobe lesion and a Wernicke's aphasia profile. This variability in recovery would seem less feasible if both left and right temporal lobes contributed equally to prelexical speech perception.

A second problem with this perspective is that is assumes the prelexical speech perception is a meaningful stage of the perceptual processing of speech. Not only is prelexical processing a term with a variety of meanings (S. K. Scott & Wise, 2004), it is used by (e.g. by Hickok and Poeppel 2007) to imply a stage of perceptual processing which is independent of, and occurs prior to, lexical processing. This interpretation may not be correct: several studies have shown that even early auditory responses to speech and sound are modified by higher-order linguistic structure (Jacquemot et al., 2003; S. K. Scott, Rosen, Lang, & Wise, 2006). In other words, the perceptual responses to speech will be unlikely to be insensitive to the postlexical information in the signal. Predictive coding models of perception (e.g. Friston & Kiebel, 2009) include these kinds of sensitivities in active perceptual processing. 'Prelexical' processing, in others words, may make a lot of sense when we discuss the kinds of information available in spoken language, but not be a discrete separate stage of post-auditory, pre-lexical processing in the human brain.

What do we want from a theory of hemispheric asymmetries in auditory processing?—A satisfactory theory of hemispheric asymmetries in auditory processing would ideally predict auditory preferences that were associated with the hemispheric preferences being accounted for: in other words, a theory that rapid temporal processing is critical to speech, should be able to demonstrate a left hemisphere preference for rapid auditory cues that has commonalties with the left hemispheric dominance in speech perception. The AST and Zatorre theories of hemispheric asymmetries in auditory processing both predict that such left hemisphere preferences should be found: however these have been empirically hard to demonstrate. In contrast, the longer time windows predicted to be processed in the right temporal lobe are in line with the data (e.g. Boemio et al., 2005), and a generous reinterpretation of the 'spectral' sense of the Zatorre theory as associated with pitch is also in line with the preferential processing of pitch variation in the right temporal lobe. While generating plausible hypotheses that the right temporal lobe is interested in longer sounds, that have pitch variation, this does not lead us to an understanding of why speech perception is associated primarily with the left temporal lobe.

It is also arguable that both theories have set the bar too high for themselves: in addition to the unquestioning adoption of the concept that rapid temporal lobe processing is critical to speech, there is no logical necessity for any acoustic sensitivity that might underlie speech perception to be mirrored by complementary processes in the right hemisphere. Other theories of hemispheric asymmetries (e.g. in cognition, visual perception, handedness, attention etc.) do not make this assumption, and we in the auditory neurosciences may be setting an unrealistic goal in trying to define dimensions of auditory sensitivities which are differentially expressed in each hemisphere.

Any theory of the auditory preferences which underlie speech perception in the left temporal lobe will also need to tackle the fact that spectral processing is critical to speech perception. A certain amount of spectrotemporal processing is needed for speech to be intelligible, and it is imprecise to suggest, as Poeppel and Giraud do in a recent review (Giraud & Poeppel, 2012), that temporal degradation of the speech signal causes "marked failures of perception", while spectral degradation can be tolerated. For example, both varying the number of channels in noise vocoded speech, and varying the smoothness of the amplitude envelope within each channel has an effect on the intelligibility of the speech, but the spectral detail dominates (Obleser et al., 2008; Shannon et al., 1995). It is time to forget about time? Certainly not. Sounds only have meaning as they evolve over time, and temporal properties of sounds (e.g. the amplitude envelope) is essential to the ways that information and meaning are expressed in sound. However we need to keep the temporal changes in the spectral domain in centre view, and this probably means using a spectrotemproal framework to think about the speech signal and to consider how this is processed neurally.

Only language left?

Although this paper has focused on approaches to speech perception, it is worth emphasizing that many aspects of language are left lateralized, whether or not, the input is from the spoken word. Thus reading speech reading and sign language perception are all left lateralized phenomena, and we have encountered no model of these processes that hypothesizes complementary left/right hemisphere asymmetries of domain general perceptual sensitivities which might underlie this. Instead, language appears to be a genuinely left hemisphere phenomenon. It remains moot, however, whether we need domain general or domain specific mechanism(s) to explain this linguistic lateralization. For example, Tim Shallice has argued that the left hemisphere is specialized for the formation of categories, and this might underlie a left hemisphere dominance for language (Tim Shallice, personal communication). Such an approach might also explain why expertise in dealing with auditory stimuli (e.g. Leech, Holt, Devlin, & Dick, 2009; Schulze, Mueller, & Koelsch, 2011) is associated with left temporal lobe regions, in the absence of overt linguistic task demands, if one assumes that perceptual expertise is associated with the development of categorical structure in perceptual experience. Another argument, from the visual domain, is that attentional processes have different characteristics in the left and right hemispheres (Brown & Kosslyn, 1995), and this may have important consequences for how auditory stimuli and speech sounds are processed – for example, there is some evidence for greater responses to unattended speech and sounds in the right temporal lobe than the left (S. K. Scott, Rosen, Beaman, Davis, & Wise, 2009).

Finally, as we discussed elsewhere (C. McGettigan & S. K. Scott, 2012), just because people have been looking for auditory preferences in the wrong places does not mean that auditory preferences are not out there to be found. We have suggested correlated sensitivities to the ways that human articulators make sounds as one candidate kind of complex auditory preferences which need not be linguistic. Other candidate processes include auditory attention, and how this interacts with the distinctly different ways that we process linguistic sounds. If a specialization for distinct acoustic properties underlie speech perception, these

properties will not be simple, fixed nor are they likely to be singular. No one cue underlies the intelligibility of human speech, and when listeners are deprived of one cue they will use another. Human speech perception rests on complex, parallel and highly plastic perceptual processes, and any auditory precursors to this will need to share similar properties of complexity and plasticity.

References

Abrams DA, Nicol T, Zecker S, Kraus N. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. Journal of Neuroscience. 2008; 28(15):3958–3965. doi: 10.1523/jneurosci.0187-08.2008. [PubMed: 18400895]

Agnew ZK, McGettigan C, Scott SK. Discriminating between Auditory and Motor Cortical Responses to Speech and Nonspeech Mouth Sounds. Journal of Cognitive Neuroscience. 2011; 23(12):4038–4047. [PubMed: 21812557]

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98(23):13367–13372. doi: 10.1073/pnas.201400998. [PubMed: 11698688]

Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A. Lateralization of speech and auditory temporal processing. Journal of Cognitive Neuroscience. 1998; 10(4):536–540. doi: 10.1162/089892998562834. [PubMed: 9712682]

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci. 2005; 8(3):389–395. doi: 10.1038/nn1409. [PubMed: 15723061]

Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD. Bihemispheric foundations for human speech comprehension. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(40):17439–17444. doi: 10.1073/pnas.1000531107. [PubMed: 20855587]

Broca P. Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletin de la Société Anatomique. 1861; 6:330–357.

Brown, HD.; Kosslyn, SM. Hemispheric differences in visual object processing: Structural versus aloocation theories. In: Davidson, RJ.; Hugdahl, K., editors. Brain asymmetry. MIT Press; Cambridge, MA: 1995. p. 77-97.

Crinion J, Price CJ. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. [Article]. Brain. 2005; 128:2858–2871. doi: 10.1093/brain/awh659. [PubMed: 16234297]

Cummins F. Oscillators and syllables: a cautionary note. Frontiers in psychology. 2012; 3:364–364. doi: 10.3389/fpsyg.2012.00364. [PubMed: 23060833]

Cutting JE. 2 Left-Hemisphere Mechanisms in Speech-Perception. Perception & Psychophysics. 1974; 16(3):601–612. doi: 10.3758/bf03198592.

D'Ausilio A, Bufalari I, Salmas P, Busan P, Fadiga L. Vocal pitch discrimination in the motor system. [Research Support, Non-U.S. Gov't]. Brain Lang. 2011; 118(1-2):9–14. doi: 10.1016/j.bandl.2011.02.007. [PubMed: 21458056]

Efron R. Temporal perception, aphasia and deja vu. Brain. 1963; 86(3):403–&. doi: 10.1093/brain/86.3.403. [PubMed: 14063892]

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations. Journal of Neuroscience. 2010; 30(21):7179–7186. doi: 10.1523/jneurosci.4040-09.2010. [PubMed: 20505085]

Fiez JA, Raichle ME, Miezin FM, Petersen SE, Tallal P, Katz WF. PET studies of auditory and phonological processing - effects of stimulus characteristics and task demands. Journal of Cognitive Neuroscience. 1995; 7(3):357–375. doi: 10.1162/jocn.1995.7.3.357. [PubMed: 23961866]

Friston K, Kiebel S. Predictive coding under the free-energy principle. [Research Support, Non-U.S. Gov't]. Philos Trans R Soc Lond B Biol Sci. 2009; 364(1521):1211–1221. doi: 10.1098/rstb. 2008.0300. [PubMed: 19528002]

Ghitza O. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Frontiers in psychology. 2011; 2:130–130. [PubMed: 21743809]

Ghitza O. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Frontiers in psychology. 2012; 3:238–238. [PubMed: 22811672]

Ghitza O. The theta-syllable: a unit of speech information defined by cortical function. [Perspective]. Frontiers in psychology. 2013; 4 doi: 10.3389/fpsyg.2013.00138.

Ghitza O, Giraud A-L, Poeppel D. Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. Frontiers in Human Neuroscience. 2013; 6 doi: 10.3389/fnhum. 2012.00340.

Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 2007; 56(6):1127–1134. doi: 10.1016/j.neuron.2007.09.038. [PubMed: 18093532]

Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience. 2012; 15(4):511–517. doi: 10.1038/nn.3063.

Griffiths TD. Central auditory pathologies. [Article]. British Medical Bulletin. 2002; 63:107–120. doi: 10.1093/bmb/63.1.107. [PubMed: 12324387]

Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: A neuropsychological analysis. Neuropsychologia. 2010; 48(4):1104–1114. doi: 10.1016/j.neuropsychologia.2009.12.011. [PubMed: 20006628]

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 2000; 4(4):131–138. [PubMed: 10740277]

Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92(1-2):67–99. [PubMed: 15037127]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8(5): 393–402. [PubMed: 17431404]

Hickok G, Poeppel D. Opinion - The cortical organization of speech processing. Nature Reviews Neuroscience. 2007; 8(5):393–402. doi: 10.1038/nrn2113.

Howard MF, Poeppel D. Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. Journal of Neurophysiology. 2010; 104(5):2500–2511. doi: 10.1152/jn.00251.2010. [PubMed: 20484530]

Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E. Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. Journal of Neuroscience. 2003; 23(29): 9541–9546. [PubMed: 14573533]

Jamison HL, Watkins KE, Bishop DVM, Matthews PM. Hemispheric specialization for processing auditory nonspeech stimuli. Cerebral Cortex. 2006; 16(9):1266–1275. doi: 10.1093/cercor/bhj068. [PubMed: 16280465]

Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain. 2000; 123:155–163. doi: 10.1093/brain/123.1.155. [PubMed: 10611129]

Johnsrude IS, Zatorre RJ, Milner BA, Evans AC. Left-hemisphere specialization for the processing of acoustic transients. Neuroreport. 1997; 8(7):1761–1765. [PubMed: 9189928]

Kelly, J.; Local, JK. Doing phonology: observing, recording, interpreting. Manchester University Press; UK: 1989.

Kimura D. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology. 1961; 15(3):166–171.

Leech R, Holt LL, Devlin JT, Dick F. Expertise with Artificial Nonspeech Sounds Recruits Speech-Sensitive Cortical Regions. Journal of Neuroscience. 2009; 29(16):5234–5239. doi: 10.1523/ jneurosci.5758-08.2009. [PubMed: 19386919]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cerebral Cortex. 2005; 15(10):1621–1631. doi: 10.1093/cercor/bhi040. [PubMed: 15703256]

Lisman JE, Jensen O. The Theta-Gamma Neural Code. Neuron. 2013; 77(6):1002–1016. [PubMed: 23522038]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54(6):1001–1010. doi: 10.1016/j.neuron.2007.06.004. [PubMed: 17582338]

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK. An application of univariate and multivariate approaches in FMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. J Cogn Neurosci. 2012; 24(3):636–652. doi: 10.1162/jocn_a_00161. [PubMed: 22066589]

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK. Speech comprehension aided by multiple modalities: Behavioural and neural interactions. Neuropsychologia. 2012; 50(5): 762–776. doi: 10.1016/j.neuropsychologia.2012.01.010. [PubMed: 22266262]

McGettigan C, Scott SK. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends Cogn Sci. 2012; 16(5):269–276. doi: 10.1016/j.tics.2012.04.006. [PubMed: 22521208]

McGettigan C, Scott SK. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends in Cognitive Sciences. 2012; 16(5):269–276. doi: 10.1016/j.tics.2012.04.006. [PubMed: 22521208]

Millman RE, Woods WP, Quinlan PT. Functional asymmetries in the representation of noise-vocoded speech. Neuroimage. 2011; 54(3):2364–2373. doi: 10.1016/j.neuroimage.2010.10.005. [PubMed: 20946961]

Morillon B, Liegeois-Chauvel C, Arnal LH, Benar C-G, Giraud A-L. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. Frontiers in psychology. 2012; 3:248–248. [PubMed: 22833730]

Mummery CJ, Ashburner J, Scott SK, Wise RJS. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J Acoust Soc Am. 1999; 106(1):449–457. doi: 10.1121/1.427068. [PubMed: 10420635]

Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex. 2003; 13(12):1362–1368. [PubMed: 14615301]

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Brugge JF. Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. Journal of Neuroscience. 2009; 29(49):15564–15574. doi: 10.1523/jneurosci.3065-09.2009. [PubMed: 20007480]

Obleser J, Eisner F, Kotz SA. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. Journal of Neuroscience. 2008; 28(32):8116–8123. doi: 10.1523/jneurosci.1290-08.2008. [PubMed: 18685036]

Obleser J, Herrmann B, Henry MJ. Neural oscillations in speech: don't be enslaved by the envelope. Frontiers in Human Neuroscience. 2012; 6 doi: 10.3389/fnhum.2012.00250.

Obleser J, Weisz N. Suppressed Alpha Oscillations Predict Intelligibility of Speech and its Acoustic Details. Cerebral Cortex. 2012; 22(11):2466–2477. doi: 10.1093/cercor/bhr325. [PubMed: 22100354]

Overath T, Kumar S, von Kriegstein K, Griffiths TD. Encoding of Spectral Correlation over Time in Auditory Cortex. Journal of Neuroscience. 2008; 28(49):13268–13273. doi: 10.1523/jneurosci.4596-08.2008. [PubMed: 19052218]

Peelle JE. The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Frontiers in Human Neuroscience. 2012; 6 doi: 10.3389/fnhum.2012.00309.

Peelle JE, Davis MH. Neural Oscillations Carry Speech Rhythm through to Comprehension. Frontiers in psychology. 2012; 3:320–320. [PubMed: 22973251]

Peelle JE, Gross J, Davis MH. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. Cerebral Cortex. 2012 doi: doi:10.1093/cercor/bhs118.

Pena M, Melloni L. Brain Oscillations during Spoken Sentence Processing. Journal of Cognitive Neuroscience. 2012; 24(5):1149–1164. [PubMed: 21981666]

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication. 2003; 41(1):245–255. doi: Doi 10.1016/S0167-6393(02)00107-3.

Praamstra P, Hagoort P, Maassen B, Crul T. Word deafness and auditory cortical function: A case history and hypothesis. Brain. 1991; 114:1197–1225. [PubMed: 2065246]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience. 2009; 12(6):718–724. doi: 10.1038/nn.2331.

Rosen S, Wise RJS, Chadha S, Conway E-J, Scott SK. Hemispheric asymmetries in speech perception: Sense, nonsense and modulations. PloS One. 2011; 6(9) 10.1371/journal.pone.0024672. doi: 10.1371/journal.pone.0024672.

Ross ED. THE APROSODIAS - FUNCTIONAL-ANATOMIC ORGANIZATION OF THE AFFECTIVE COMPONENTS OF LANGUAGE IN THE RIGHT-HEMISPHERE. Archives of Neurology. 1981; 38(9):561–569. [PubMed: 7271534]

Saygin AP, Dick F, Wilson SW, Dronkers NF, Bates E. Neural resources for processing language and environmental sounds - Evidence from aphasia. [Article]. Brain. 2003; 126:928–945. doi: 10.1093/brain/awg082. [PubMed: 12615649]

Schonwiesner M, Rubsamen R, von Cramon DY. Hemispheric asymmetry for spectral and temporal processing in the human antero lateral auditory belt cortex. European Journal of Neuroscience. 2005; 22(6):1521–1528. doi: 10.1111/j.1460-9568.2005.04315.x. [PubMed: 16190905]

Schulze K, Mueller K, Koelsch S. Neural correlates of strategy use during auditory working memory in musicians and non-musicians. European Journal of Neuroscience. 2011; 33(1):189–196. doi: 10.1111/j.1460-9568.2010.07470.x. [PubMed: 21073548]

Schwartz J, Tallal P. Rate of Acoustic Change May Underlie Hemispheric-Specialization for Speech-Perception. Science. 1980; 207(4437):1380–1381. [PubMed: 7355297]

Scott S, McGettigan C. Amplitude onsets and spectral energy in perceptual experience. Front Psychol. 2012; 3:80. doi: 10.3389/fpsyg.2012.00080. [PubMed: 22470359]

Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000; 123:2400–2406. [PubMed: 11099443]

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 2003; 26(2):100–107. doi: 10.1016/s0166-2236(02)00037-1. [PubMed: 12536133]

Scott SK, Rosen S, Beaman CP, Davis JP, Wise RJS. The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. J Acoust Soc Am. 2009; 125(3):1737–1743. doi: 10.1121/1.3050255. [PubMed: 19275330]

Scott SK, Rosen S, Lang H, Wise RJS. Neural correlates of intelligibility in speech investigated with noise vocoded speech-A positron emission tomography study. Journal of the Acoustical Society of America. 2006; 120(2):1075–1083. doi: 10.1121/1.2216725. [PubMed: 16938993]

Scott SK, Wise RJS. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 2004; 92(1-2):13–45. doi: 10.1016/j.cognition.2002.12.002. [PubMed: 15037125]

Shankweiler D, Studdert-Kennedy M. Identification of consonants and vowels presented to left and right ears. Q J Exp Psychol. 1967; 19(1):59–63. doi: 10.1080/14640746708400069. [PubMed: 6041684]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. SPEECH RECOGNITION WITH PRIMARILY TEMPORAL CUES. Science. 1995; 270(5234):303–304. doi: 10.1126/science.270.5234.303. [PubMed: 7569981]

Studdert-Kennedy M, Mody M. Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. Psychonomic Bulletin & Review. 1995; 2(4):508–514. doi: 10.3758/bf03210986. [PubMed: 24203788]

Studdert-Kennedy M, Shankweiler D. HEMISPHERIC39 SPECIALIZATION FOR LANGUAGE PROCESSES. Science. 1981; 211(4485):960–961. doi: 10.1126/science.7466372. [PubMed: 7466372]

Tallal P, Piercy M. Defects of nonverbal auditory-perception in children with developmental aphasia. Nature. 1973; 241(5390):468–469. doi: 10.1038/241468a0. [PubMed: 4705758]

Tallal P, Piercy M. DEVELOPMENTAL APHASIA - RATE OF AUDITORY PROCESSING AND SELECTIVE IMPAIRMENT OF CONSONANT PERCEPTION. Neuropsychologia. 1974; 12(1): 83–93. doi: 10.1016/0028-3932(74)90030-x. [PubMed: 4821193]

Tallal P, Stark RE. SPEECH ACOUSTIC-CUE DISCRIMINATION ABILITIES OF NORMALLY DEVELOPING AND LANGUAGE-IMPAIRED CHILDREN. Journal of the Acoustical Society of America. 1981; 69(2):568–574. doi: 10.1121/1.385431. [PubMed: 7462479]

Van der Haegen L, Westerhausen R, Hugdahl K, Brysbaert M. Speech dominance is a better predictor of functional brain asymmetry than handedness: a combined fMRI word generation and behavioral dichotic listening study. Neuropsychologia. 2013; 51(1):91–97. doi: 10.1016/j.neuropsychologia. 2012.11.002. [PubMed: 23149380]

Wernicke, C. Der aphasische Symptomencomplex. Cohn & Weigert; Breslau: 1874.

Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak R. Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain. 1991; 114:1803–1817. [PubMed: 1884179]

Zaehle T, Wustenberg T, Meyer M, Jancke L. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. European Journal of Neuroscience. 2004; 20(9):2447–2456. doi: 10.1111/j.1460-9568.2004.03687.x. [PubMed: 15525285]

Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. [Research Support, Non-U.S. Gov't]. Cereb Cortex. 2001; 11(10):946–953. doi: 10.1093/cercor/11.10.946. [PubMed: 11549617]

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences. 2002; 6(1):37–46. doi: 10.1016/s1364-6613(00)01816-7. [PubMed: 11849614]

Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. Philosophical Transactions of the Royal Society B-Biological Sciences. 2008; 363(1493):1087–1104. doi: 10.1098/rstb.2007.2161.
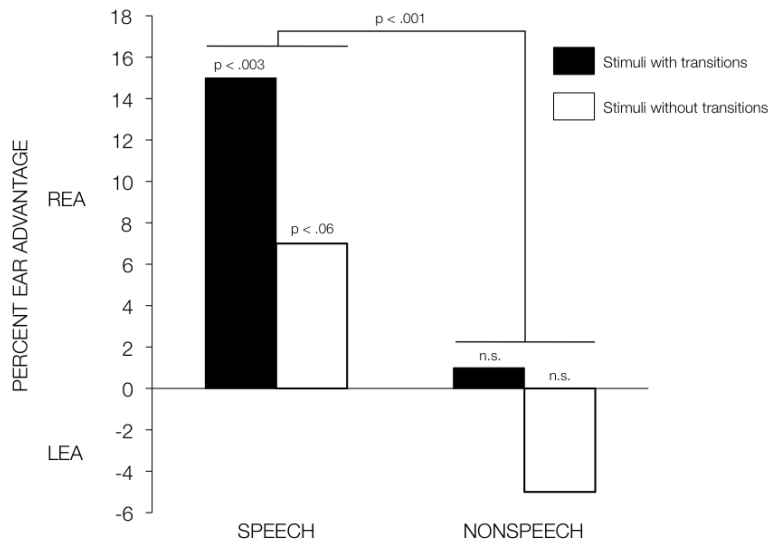
Figure 1.
shows the right/left ear advantage in the processing of synthetic consonant vowel and vowel stimuli, and of sine wave (non speech) equivalents of these stimuli, replotted from the data presented in Cutting (1974).