



Comprehending auditory speech: previous and potential contributions of functional MRI

Samuel Evans & Carolyn McGettigan

To cite this article: Samuel Evans & Carolyn McGettigan (2017) Comprehending auditory speech: previous and potential contributions of functional MRI, *Language, Cognition and Neuroscience*, 32:7, 829-846, DOI: [10.1080/23273798.2016.1272703](https://doi.org/10.1080/23273798.2016.1272703)

To link to this article: <https://doi.org/10.1080/23273798.2016.1272703>



Published online: 09 Jan 2017.



Submit your article to this journal [↗](#)



Article views: 451



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

REVIEW ARTICLE



Comprehending auditory speech: previous and potential contributions of functional MRI

Samuel Evans^{a*} and Carolyn McGettigan^{a,b}

^aInstitute of Cognitive Neuroscience, University College London, London, UK; ^bDepartment of Psychology, Royal Holloway, University of London, London, UK

ABSTRACT

Functional neuroimaging revolutionised the study of human language in the late twentieth century, allowing researchers to investigate its underlying cognitive processes in the intact brain. Here, we review how functional MRI (fMRI) in particular has contributed to our understanding of speech comprehension, with a focus on studies of intelligibility. We highlight the use of carefully controlled acoustic stimuli to reveal the underlying hierarchical organisation of speech processing systems and cortical (a)symmetries, and discuss the contributions of novel design and analysis techniques to the contextualisation of perisylvian regions within wider speech processing networks. Within this, we outline the methodological challenges of fMRI as a technique for investigating speech and describe the innovations that have overcome or mitigated these difficulties. Focussing on multivariate approaches to fMRI, we highlight how these techniques have allowed both local neural representations and broader scale brain systems to be described.

ARTICLE HISTORY

Received 25 January 2016
Accepted 4 November 2016

KEYWORDS

Speech; language; brain;
fMRI; perception

General introduction

The emergence of functional neuroimaging in the 1990s offered the unprecedented opportunity to examine the processing of spoken language in the intact, functioning human brain with a spatial resolution on the order of millimetres. Building upon the highly influential neuropsychological literature of the late nineteenth and twentieth centuries, a key aim of empirical study was uncovering the neural substrates for the comprehension of speech. Here, we offer a review and synthesis of the contributions of functional imaging – in particular, functional magnetic resonance imaging (fMRI) – to our understanding of the functional neuroanatomy of speech comprehension. We also offer an evaluation of the method and ask whether fMRI will continue to contribute substantially to our knowledge, or whether other modalities (or combinations of techniques) hold more promise for the future of the field.

Other authors have offered excellent overviews of the state of the art in the functional imaging of language processes more generally (e.g. Price, 2012), and in the specific methodological challenges of examining auditory processing with fMRI (Peelle, 2014). For this review, we felt that it was timely to limit our discussion to the use of fMRI in the investigation of speech comprehension. The study of speech comprehension has been at the centre of

significant theoretical debate in the last decade, for example, in addressing the role of hemispheric asymmetries and motor contributions to speech perception. Advances in experimental design and data analysis have played an important role in refining our understanding of these issues. Here, we highlight in particular the contributions and future opportunities afforded by multivariate analyses of the blood oxygenation level dependent (BOLD) response in examining the nature and content of neural representations of the speech signal. Approaches such as representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008), provide promise for integrating multiple imaging modalities, allowing examination of neural responses at high levels of spatial and temporal resolution (e.g. EEG and fMRI). Further, multivariate network-based analyses such as independent components analysis allow for the disentangling of overlapping neural processes in speech comprehension, such that speech-specific processes can be dissociated from those associated with task difficulty and domain-general functions (see Adank, 2012; Brownsett et al., 2014; Geranmayeh, Brownsett, & Wise, 2014).

For the purposes of this short article, we define speech comprehension as the recognition of intelligible phonemes, syllables, words and sentences, where we are namely concerned with the processes supporting

CONTACT Carolyn McGettigan  carolyn.mcgettigan@rhul.ac.uk

*Present address: Department of Psychology, University of Westminster, 115 New Cavendish Street, London, W1W 6UW, UK

© 2017 Informa UK Limited, trading as Taylor & Francis Group

the mapping of auditory percepts onto stored linguistic representations. In reviewing the literature, we acknowledge the tension between the preferential study of naturalistic linguistic stimuli (i.e. sentences and passages, vs. words and syllables) with the observation that such stimuli engage brain regions involved in higher level cognitive processes (both domain-general and domain-specific) beyond the basic extraction of an intelligible percept (reviewed in Price, 2012). We do not have the scope for an exhaustive overview of studies in the field and thus concentrate our focus on comprehension at the word and sentence level. However, we note that any complete account of speech comprehension must crucially be accompanied by a solid understanding of the earlier stages in the speech processing hierarchy (Oblaser & Eisner, 2009) – in later sections of our discussion, we highlight how emerging multivariate approaches to fMRI data will enhance descriptions of pre-lexical processing beyond the current state of the art.

In the sections that immediately follow (and predominantly for readers unfamiliar with the field), we offer a contextual introduction to the study of speech comprehension using functional imaging of the healthy brain, and a brief examination of the key methodological issues in fMRI (and how these have been addressed). The main thrust of our discussion then considers the theoretical developments in the field, and how these are being advanced through recent methodological innovations. We concentrate on three main theoretical issues through which fMRI of healthy young brains has provided key empirical advances: investigation of anterior versus posterior temporal lobe contributions to comprehension, hemispheric asymmetries, and the contribution of wider brain networks to comprehension processes. We add to these some additional consideration of how fMRI studies of other populations – namely, healthy older listeners and those with acquired language impairments – have helped to inform our understanding from the perspective of cortical reorganisation and compensation. Finally, we close our thoughts by considering how fMRI can continue to advance the field through continued refinement of the method and greater integration with other imaging and measurement modalities.

Historical context: the emergence of fMRI over PET in the study of speech

There was a period of time during the 1990s when positron emission tomography (PET) and fMRI co-existed as similarly popular methods for the investigation of speech comprehension, and several seminal studies emerged from studies employing these methods (Price, 2012). However, fMRI offered several substantial

methodological advantages over PET. First, and crucially, the non-invasive nature of fMRI (cf. the injection of labelled tracers for PET) permitted scanning of a more representative section of the population (i.e. to include children and women of child-bearing age), and more easily allowed for repeated scanning of the same subjects. Further, the ability to collect whole-brain functional images in just a few seconds allowed neuroimaging researchers to adopt behavioural paradigms with larger numbers of trials that were better aligned with the randomised, event-related designs used in experimental psychology. However, the production of intense noise due to gradient switching in the magnetic coils of MRI scanners (see “Challenges and Solutions” below) presented a particular challenge for the study of auditory perception that formed a distinct and limiting disadvantage compared with the relative quiet of PET scans.

Our reading of the literature from the end of the twentieth century identifies the most significant advances in understanding as arising from the use of more sophisticated acoustic baseline conditions in subtraction analyses that compared the perception of speech with unintelligible sounds (for discussions of cognitive subtraction, see Caplan, 2009; Friston et al., 1996). Scott, Blank, Rosen, and Wise (2000) reported a passive listening study in PET, employing two forms of intelligible speech (natural recordings and noise-vocoded transformations) and two spectrally rotated unintelligible equivalents, matched in acoustic complexity. Regions in posterior superior temporal gyrus (STG) and superior temporal sulcus (STS) were sensitive to the presence of acoustic-phonetic content, while activation magnitude in the anterior STS was sensitive to stimulus intelligibility but not acoustic variation across the different forms of intelligible speech. With both activations expressed dominantly in the left hemisphere, these results were coherent with the existing aphasia literature. Further, the suggestion of an anterior-going hierarchy of speech processing in the human temporal lobe was consistent with work from non-human primates, which described a hierarchically organised ventral auditory object recognition pathway radiating from primary auditory cortex (Hackett, 2011; Kaas & Hackett, 2000; Kaas, Hackett, & Tramo, 1999; Rauschecker & Tian, 2000).

These relatively simple subtraction designs were elaborated upon by later work that took advantage of the possibility to run more complex, event-related designs in fMRI (Belin, Zatorre, Hoge, Evans, & Pike, 1999; Dale, 1999; Friston, Zarahn, Josephs, Henson, & Dale, 1999). Thus, Davis and Johnsrude (2003) employed a correlational design in which they identified brain regions whose activation increased parametrically with the increasing intelligibility of sentences that had been

degraded in different ways. They further differentiated between regions that were, and were not, additionally sensitive to the acoustic make-up of the stimulus, or whose responses were particularly enhanced for effortful listening conditions. This revealed, in line with the PET work, a hierarchical pathway of acoustic to intelligible speech processing with increasing distance from primary auditory cortex, in which both the anterior and posterior STS were sensitive to intelligibility but not to variation in acoustic form. Studies probing the abstraction of pre-lexical information in speech also showed evidence for a processing hierarchy, finding non-categorical acoustic responses to stop consonants and complex non-speech sounds in planum temporale (posterior to Heschl's gyrus), with preferential responses to speech (compared with non-speech control conditions) emerging and extending anteriorly on the STS (Oblaser et al., 2006; Oblaser, Zimmermann, Van Meter, & Rauschecker, 2007). However, in general the understanding of pre-lexical processes and representations in speech was limited by the relatively coarse spatial resolution of fMRI, and the spatial averaging employed in univariate subtraction analyses (Oblaser & Eisner, 2009). Hence, the evidence for categorical responses to phonemes in auditory cortex was limited (e.g. Oblaser et al., 2006). Later in this paper, we shall return to this issue in our discussion of more recent multivariate techniques that have to some extent addressed these limitations.

Importantly, in addition to peri-auditory activation, studies of speech comprehension reported relevant contributions from the left inferior frontal gyrus (IFG). For example, Davis and Johnsrude (2003) found preferential responses to intelligible degraded speech within the IFG that were insensitive to the acoustic form of the degradation. Elucidating the role of the IFG, Binder, Liebenthal, Possing, Medler, and Ward (2004) associated insular and frontal opercular sites with response selection, while temporal lobe activations were linked more closely with perceptual processes. Similarly, a number of studies of pre-lexical processing associated the perception of across-versus within-category phonemic contrasts with sites in inferior parietal cortex, including the supramarginal gyrus (SMG; e.g. Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Raizada & Poldrack, 2007). Later in our discussion, we address how the growing body of research pointed towards a wider speech comprehension network extending beyond the temporal lobes. We also examine more closely recent work on the elaboration of hierarchical pathways, reviewing how knowledge has progressed concerning the relative roles of posterior and anterior temporal sites, as well as the question of hemispheric dominance in the extraction of intelligible messages from spoken language.

Overview of the method

Functional magnetic resonance imaging, similarly to PET, operates on the principle that neural metabolism is supported by the delivery of oxygen in the bloodstream. It is specifically based on the finding that deoxygenated and oxygenated haemoglobin have different properties in a magnetic field – the former paramagnetic, the latter diamagnetic (Bandettini, Wong, Hinks, Tikofsky, & Hyde, 1992; Kwong et al., 1992; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Ogawa et al., 1993). A calculation of the ratio of deoxy- to oxyhaemoglobin within a given brain voxel (i.e. a 3-dimensional pixel, typically sized 2–3 mm in each dimension and potentially containing up to 100,000 neurons) is then used as a proxy for neural activity within that portion of tissue. The haemodynamic response to an event, known as the BOLD response, is characterised by a brief decrease in signal (reflecting a local increase in metabolism), followed by an increase, peaking at 4–6 seconds after the event (accompanying an increase in local blood flow) and a slower return to baseline levels (on the order of 10–30 seconds). The precise relationship between the BOLD response and the underlying neural physiology is not fully understood, but is believed to be more strongly related to more slowly fluctuating local field potentials than to neuronal spiking activity (Logothetis & Wandell, 2004). Depending on the size and number of voxels collected, a whole-brain scan can now be acquired very quickly, typically in around 2–3 seconds. In the future, the more widespread use of multiband acquisition sequences, in which multiple brain slices are acquired simultaneously, will dramatically reduce acquisition times (see Correia, Jansma, & Bonte, 2015; De Martino, Moerel, Ugurbil, Formisano, & Yacoub, 2015) allowing greater flexibility in experimental design and wider “whole brain” coverage. These increased acquisition speeds are likely to encourage the routine collection of larger brain volumes, with the effect that anatomical regions that are often ignored, but that may still play a role in perception, are captured in the field of view more frequently (e.g. the cerebellum, see Guediche, Holt, Laurent, Lim, & Fiez, 2015; see also Visser, Jefferies, & Lambon Ralph, 2010, on the impact of field of view size on the ventral portions of the temporal lobe). In addition to faster acquisition sequences, we envisage more widespread use of higher field strength magnets; indeed, many centres have now replaced or supplemented their 1.5-Tesla systems with 3-Tesla scanners. These higher strength magnets provide improved spatial resolution, albeit at a risk of greater imaging artefact. A new generation of 7-Tesla systems is now in active use in human research, affording even higher resolution (e.g.

<1 mm in-plane; Pfeuffer et al., 2002; Yacoub et al., 2003), and work has begun in applying this greater resolution to the auditory system (Formisano et al., 2003).

Challenges and solutions

In some ways, it seems unlikely that fMRI would have superseded PET as the imaging modality of choice for researchers interested in localising speech comprehension networks. Indeed, fMRI scanners generate considerable acoustic noise, caused by switching of the gradient coils, which can generate levels upwards of 130 dB SPL (Foster, Hall, Summerfield, Palmer, & Bowtell, 2000). This noise is periodic, with harmonics in the range crucial for speech perception, and thus greatly reduces the audibility of speech signals (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999; Peelle, 2014). The acoustic properties of the scanner noise can also interact with those of the auditory stimuli, causing some stimuli or experimental conditions to be more affected by scanner noise than others, thus distorting experimental outcomes. At a neural level, listening to speech in the context of scanner noise engages compensatory neural activity associated with modulation of attention and cognitive control. In addition, the response of the auditory cortex can become saturated, reducing the dynamic range of the response to auditory stimuli (Langers, Van Dijk, & Backes, 2005; Peelle, 2014; Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010). The challenge of acoustic noise has been addressed through the development of “sparse sampling”, where a delay of ~6–18 seconds is introduced between successive acquisitions to allow for the presentation of auditory stimuli in relative quiet (Edmister et al., 1999; Hall et al., 1999). This slow sparse sampling depends on the careful placement of auditory events such that the following brain volume captures the peak of the event-related BOLD response. More recently, fast sparse acquisition sequences have been adopted that take advantage of shorter brain volume acquisition times. In this approach, the peak response is not captured in the immediately following acquisition, but rather in the subsequent one, with overlapping neural responses deconvolved using a canonical hemodynamic response function. This requires stimuli to be presented in a much shorter silent interval between successive volumes (e.g. ~ every 2–3 seconds) facilitating more stimulus presentations and greater data acquisition per unit time (see Evans & Davis, 2015; Formisano, De Martino, Bonte, & Goebel, 2008). Further developments include the Interleaved Silent Steady State (ISSS) approach (Schwarzbauer, Davis, Rodd, & Johnsrude, 2006), which involves the sequential acquisition of several volumes with interleaved silence between clusters, permitting

greater within-trial temporal resolution of the event-related BOLD response suitable for the examination of responses unfolding over time (e.g. during sentence comprehension). Others have endeavoured to reduce the magnitude or impact of the acoustic scanner noise itself by noise cancellation or manipulating the parameters of the acquisition sequence (Hall et al., 2009; Peelle et al., 2010; Schmitter et al., 2008). For a helpful and succinct description of these approaches, we highly recommend Peelle (2014).

The degradation of the BOLD signal in regions of the brain that are located close to air–tissue interfaces is an additional problem specific to fMRI data collection. With conventional gradient echo planar imaging, “signal drop out” is generated by air–tissue interfaces that cause local differences in magnetic susceptibility, inducing geometric distortions and reducing BOLD sensitivity. Unfortunately the anterior and inferior temporal cortex are particularly affected by this issue, which is a problem as these regions have been shown to play important roles in semantic processing and operate as key nodes in the wider speech comprehension network (Devlin et al., 2000; Hickok & Poeppel, 2007; Patterson, Nestor, & Rogers, 2007; Rauschecker & Scott, 2009; Rodd, Davis, & Johnsrude, 2005). Recently, dual-echo gradient imaging has been applied to address this issue. Dual-echo imaging uses two echoes, rather than a single long echo time typically used in gradient echo imaging: one echo is short, reducing the effect of field inhomogeneity in areas near bone–tissue interfaces but with a reduced BOLD sensitivity in other regions, and the other is long, producing the converse sensitivity. By combining echoes, either by averaging or taking a weighted sum, it is possible to balance sensitivities to gain acceptable signal quality across the whole brain (Poser, Versluis, Hoogduin, & Norris, 2006). This method has recently been shown to out-perform single spin-echo imaging, and has also been shown to ameliorate “signal drop out”, in temporal lobe regions (Halai, Welbourne, Embleton, & Parkes, 2014). Indeed, using this approach a recent study revealed selective responses to intelligible speech in inferior regions of the anterior temporal lobe, as well as more superior regions typically identified in studies of speech intelligibility (Halai, Parkes, & Welbourne, 2015).

An important distinction between auditory and visual processing in the human brain is the larger number of key relay nuclei in the ascending auditory pathway, including the cochlear nucleus, the superior olive, inferior colliculus and medial geniculate nucleus. Imaging these small structures presents challenges in functional imaging through the increased movement artefact associated with the proximity of these sites to

pulsatile movement in blood vessels. In the past, this artefact may have acted to obscure subcortical contributions to speech comprehension. One solution to this issue has been to employ cardiac “gating”, in which acquisition is synchronised to a point in the cardiac cycle to reduce the effect of physiological movement on the MR signal. Indeed, using cardiac gating Erb, Henry, Eisner, and Obleser (2013) showed that during adaptation to noise-vocoded speech, activity within the thalamus was down-regulated over time, suggestive that cortico–striato–thalamic loops act to sharpen representations of degraded speech during initial exposure.

The timecourse of the development and widespread adoption of these compensatory methods for imaging speech comprehension is likely to have impacted on the field, because different authors’ accounts of the speech comprehension system have been based on varied methodological approaches. In a recent meta-analysis of methodological choices in the study of speech comprehension, Adank (2012) reports, for example, that studies employing sparse sampling techniques have revealed more extensive activations in STS than those using continuous acquisition, while continuous protocols have more strongly engaged regions in the frontal-opercular control network such as the supplementary motor area (SMA) and the anterior cingulate gyrus. In related work on functional imaging studies of semantic comprehension, it was revealed that the likelihood of identifying anterior temporal lobe activation was negatively impacted by a number of factors, including the use of fMRI versus PET, and by the use of an insufficient field of view to capture tissue in the anterior temporal lobe (ATL) region (Visser et al., 2010). These factors bear implications for key questions in the literature, for example regarding the relative contributions of anterior versus posterior temporal lobe regions to speech comprehension (see “Key empirical contributions”).

Key empirical contributions

In reviewing the literature, and in particular the work of the last 10 years, we have identified several key empirical issues that have been addressed with fMRI in the study of speech comprehension. Below, we describe each in turn, and evaluate the extent to which the existing work has advanced our understanding:

Understanding the temporal hierarchy: anterior versus posterior responses to intelligible signals

The seminal study by Scott et al. (2000) presented a strong argument that the anterior STS formed the apex of the speech processing hierarchy in the superior

temporal lobes. However, while other studies identified peak responses to intelligibility in anterior temporal sites, in many of these cases activations with similar profiles were also found in the posterior STS (Davis & Johnsrude, 2003; Narain et al., 2003). In some cases, intelligibility contrasts yielded only posterior activations within the STS (e.g. Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010). Adank’s (2012) meta-analysis indicated that some of this variability may have been due to design and protocol choices across different studies. However, the development of newer methods for sensitive multivariate and connectivity analyses also allowed researchers to address this issue empirically.

In the last 15 years, multivariate pattern analysis (MVPA) techniques, adopted from the field of machine learning, have been increasingly applied to fMRI data (for tutorial reviews, see Haynes & Rees, 2006; Kriegeskorte et al., 2008; Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006; O’Toole et al., 2007; Pereira, Mitchell, & Botvinick, 2009). Traditional univariate approaches based on general linear modelling identify whether the average neural activation to one experimental condition is greater than to another, at each individual voxel. By contrast, multivariate methods consider the spatial pattern of activation across multiple voxels. This allows weakly discriminative information that is distributed over multiple voxels to be pooled (Haynes & Rees, 2006); and the representational geometry of neural responses to be described (Kriegeskorte et al., 2008; Kriegeskorte & Kievit, 2013). This can permit detection of subtle effects not observed when considering the averaged magnitude of response within isolated voxels (Bonte, Hausfeld, Scharke, Valente, & Formisano, 2014; Formisano et al., 2008; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009). However, whilst MVPA methods can sometimes afford greater sensitivity over univariate methods, this sensitivity can also lead to greater vulnerability to false positives if additional care is not taken in experimental design and analysis (see Mumford, Davis, & Poldrack, 2014; Pereira et al., 2009; Todd, Nystrom, & Cohen, 2013).

Typically in MVPA, algorithms are used to learn a rule that correctly separates brain images belonging to different experimental conditions. Different algorithms apply different criteria for making this separation. Support vector machines (SVMs) have proved popular because they have high accuracy and perform well with large numbers of voxels (see Misaki, Kim, Bandettini, and Kriegeskorte (2010) for a comparison of algorithms). An SVM is a discriminant classifier that learns a separating boundary that maximises the distance between the most similar examples from each experimental condition within a multidimensional space with as many

dimensions as voxels. The success of the boundary is tested by assessing how well it predicts category membership for independent test data, which are usually taken from a different acquisition run. If the boundary predicts category membership at a level greater than chance, this provides evidence that the brain images contain information capable of distinguishing between the conditions. In addition to choosing an algorithm, researchers must also decide which voxels to include in classification, as classifiers can perform poorly with too many voxels. Voxels can be selected using data driven approaches, such as recursive feature elimination that removes uninformative voxels (Formisano et al., 2008) or using functionally or anatomically defined regions (Evans & Davis, 2015). Alternatively, local information can also be mapped across the whole brain by extracting patterns from small cortical patches iteratively, referred to as a searchlight procedure (Kriegeskorte, Goebel, & Bandettini, 2006). Whole brain volumes can also be classified to understand how multivariate information is integrated at a larger scale, using dimension reduction techniques such as principal component analysis (Mourao-Miranda, Bokde, Born, Hampel, & Stetter, 2005).

In a fMRI study of sentence comprehension using the same auditory conditions as Scott et al. (2000), Okada et al. (2010) used an SVM to measure the discriminability of multivoxel response patterns to pairs of conditions, in a set of regions of interest (ROIs) throughout the left and right superior temporal cortex. They used classification scores to calculate an “acoustic invariance index” that expressed intelligibility classifications (e.g. accuracy in discriminating responses to intelligible non-rotated sentences from unintelligible spectrally rotated versions) relative to spectral detail classifications (e.g. noise vocoded from non-vocoded sentences). Using this metric, they argued that posterior STS showed the greatest invariance to the acoustic properties of the stimuli. In another fMRI replication of this paradigm, Evans, Kyong et al. (2014) used both univariate and multivariate analyses, showing the strongest univariate intelligibility response in left anterior STS but the greatest multivariate classification accuracy in local patches of cortex (searchlights) in the posterior STS. Yet, when exploring classifications in which information could be integrated across the whole temporal lobe, they found that the most informative voxels showing preferential responses to intelligible speech were located in anterior regions. Taken together, these findings suggest that both anterior and posterior STS occupy higher, more abstract levels of processing and that there might be multiple hierarchies within which information can be encoded at different spatial scales.

By now, a number of studies have used pattern classification as a metric for quantifying the relative contribution of different regions to acoustic and linguistic processes in the case of connected speech signals (Abrams et al., 2013; Evans et al., 2014; McGettigan et al., 2012; Okada et al., 2010). Correia et al. (2014) extended this approach using an innovative cross decoding approach (see Kaplan, Man, & Greening, 2015, for a review of cross decoding) to demonstrate the encoding of abstract word representations independent of speech acoustics. In the study, they presented single words in Dutch and English to Dutch–English bilinguals. They trained an SVM to distinguish between neural patterns elicited by words presented in Dutch, and applied this learned boundary to predict the classification of the same words presented in English (and vice versa). A number of regions showed accurate cross decoding, including the left anterior and right posterior temporal cortex, and regions of the parietal, frontal and occipital cortices, suggesting a common semantic representation for individual words in the absence of a shared acoustic signal. Successful classification in the anterior temporal lobe in this study, in the absence of the use of distortion-corrected EPI, also highlights the sensitivity of MVPA to detect effects in regions of low signal to noise ratio, such as those affected by susceptibility artefact (Kriegeskorte & Bandettini, 2007).

Obleser and Eisner (2009) were amongst the first to highlight the opportunities afforded by MVPA to investigate how information is coded at finer, pre-lexical, levels of processing. This approach has indeed led to important insights concerning how speech sounds are represented in the human brain (Arsenault & Buchsbaum, 2015; Du, Buchsbaum, Grady, & Alain, 2014; Formisano et al., 2008; Lee, Turkeltaub, Granger, & Raizada, 2012; Obleser, Leaver, Vanmeter, & Rauschecker, 2010; Raizada, Tsao, Liu, & Kuhl, 2010). In the first study to apply MVPA techniques to speech perception data, Formisano et al. (2008) trained a classifier to discriminate between vowels from one speaker, and applied this learned boundary to discriminate between the same vowels spoken by another speaker. This provided evidence for the instantiation of speech sound categories invariant to speaker identity. Using recursive feature elimination, they identified that decoding of vowels could be achieved using a distributed set of voxels spreading across the STG and into anterior lateral Heschl’s Gyrus. A later study replicated and extended this finding, showing that task demands – that is, whether participants attended to the identity of the vowel rather than the speaker – selectively enhanced discriminability of the activity patterns in superior temporal cortex, which extended “backwards” into early auditory

regions (Bonte et al., 2014). In agreement, Obleser et al. (2010) used an SVM to discriminate between a set of CV syllables spoken by multiple speakers that could be classified either by their vowel (front or back vowel) or consonant category (front or back stop consonant). They also found evidence for distributed neural patterns extending anteriorly and posteriorly within the superior temporal cortex for both the vowel and consonant discriminations, with only a sparse overlap of voxels that were capable of both classifications. As in the previously described studies, voxels in and around primary auditory cortex contributed to discriminations between speech sound categories. Kilian-Hutten, Valente, Vroomen, and Formisano (2011) presented physically identical ambiguous speech sounds in which perception could be recalibrated by previous visual speech exposure. They showed that regions including the posterior bank of Heschl's sulcus and gyrus contributed to discriminating the subjective identity of the syllables when the sounds themselves were physically identical, suggesting that early auditory cortex contributes to higher-order constructive processes in perception.

Taken together, these results suggest that abstract speech sound categories are encoded in sparse, distributed activity patterns. These patterns span anterior and posterior temporal cortex and extend into early auditory regions, including Heschl's gyrus. This suggests that speech processing hierarchies may be instantiated in a less discrete and more graded manner than previously thought. Indeed this work suggests that early auditory cortex contributes to higher-order perceptual processes, rather than just passively responding to low level acoustic features. However, these data are not incompatible with a hierarchical account of perception. Indeed, a recent study by Evans and Davis (2015) demonstrated that the information contained within primary auditory cortex was not sufficient on its own to represent the categorical identity of speech syllables, suggesting that early auditory cortex contributes to perception, but only via its co-activation with downstream regions of the speech processing system.

Important insights concerning the role of posterior and anterior temporal cortex have also come from functional connectivity analyses. Early connectivity studies described co-activation between regions using simple correlations (Eisner et al., 2010; Obleser, Wise, Dresner, & Scott, 2007). More recently, *effective* connectivity techniques, such as dynamic causal modelling (DCM) and Granger causality, have been used to make stronger inferences about the directionality of influence between brain regions (for tutorial reviews, see Stephan et al. (2010) and Friston, Moran, and Seth (2013)). Using Granger causality, Upadhyay et al. (2008)

found that spoken sentences engaged two streams of processing emanating from primary auditory cortex – one to the anterior STG and the other to posterior STG – paralleling findings from studies of non-human primates indicating multiple processing streams. Leff et al. (2008) identified preferential responses to intelligible two-word phrases in left posterior and anterior STS, and pars orbitalis of the left IFG. In a DCM analysis, they found that the best model to explain the interactivity of these three sites involved auditory inputs driving posterior STS, with the intelligibility of speech modulating feed-forward connectivity from there to the other two sites. A more recent study has replicated this evidence for feed-forward connectivity between posterior and anterior sections of the STS, for both tonal and non-tonal languages (Ge et al., 2015).

In sum, the work of the last ten years, with support from emerging techniques including MVPA and DCM, indicates important contributions for both posterior and anterior temporal cortex in speech comprehension. Evidence suggests that distributed, sparse neural codes support perception, and these patterns extend across anterior and posterior temporal cortex and into early auditory regions. The anterior temporal cortex may play a later role in the speech processing hierarchy, with the possibility that information may be represented at different spatial scales in anterior as compared to posterior regions.

Cortical (a)symmetries: is speech comprehension left-dominant, or bilateral?

Scott et al. (2000) reported strongly left-dominant responses to acoustic-phonetic content and speech intelligibility in the temporal lobes, yet using PET were unable to provide any formal statistical comparison of the effects. Nonetheless, the body of work that followed tended to show larger effect sizes to intelligible speech than to complex acoustic baseline sounds in the left superior temporal cortex (Friederici, Kotz, Scott, & Obleser, 2010; Narain et al., 2003; Obleser, Wise, et al., 2007). The remaining problem, however, was how to quantify this: comparing the number of significant voxels across hemispheres is dependent on the choice of statistical threshold (although see Evans, McGettigan, Agnew, Rosen, & Scott, 2016; Kyong et al., 2014, for an alternative univariate approach), there are anatomical differences between left and right temporal lobes that limit direct comparability via subtraction analysis (though this may be addressed through normalisation to a symmetrical template brain; e.g. Herrmann, Obleser, Kalberlah, Haynes, & Friederici, 2012; Watkins et al., 2001), and use of ROI approaches is limited by

factors such as a lack of statistical independence or arbitrariness in the size and/or shape of analysis regions (McGettigan et al., 2012). Another challenge was how to tease apart acoustic from linguistic contributions to asymmetries. A large part of the literature had focused on exploring hemispheric sensitivities to basic acoustic information in sounds and relating this to the properties of speech (e.g. fast vs. slow modulations/short vs. long time windows; see Boemio, Fromm, Braun, & Poeppel, 2005; Giraud & Poeppel, 2012; Hickok & Poeppel, 2007; Poeppel, 2003), or spectral vs. temporal processing (e.g. Schonwiesner, Rubsamen, & von Cramon, 2005; Zatorre & Belin, 2001; Zatorre & Gandour, 2008; Zatorre, Belin, & Penhune, 2002). However, other authors chose to examine responses to acoustic modulations in the context of their impact on the intelligibility of connected speech signals (e.g. McGettigan et al., 2012; Obleser, Eisner, & Kotz, 2008; Rosen, Wise, Chadha, Conway, & Scott, 2011). Obleser et al. (2008) parametrically varied the number of spectro-temporal channels of information in noise-vocoded sentences, as well as the temporal smoothing of the amplitude envelope, in a fully factorial design in fMRI. They found that both manipulations affected speech comprehension ratings, and both engaged BOLD responses in bilateral superior temporal cortex. However, they showed that the laterality of neural responses to the temporal smoothing factor was more left-dominant, whereas that for the spectro-temporal channels modulation was right-ward (though note that laterality was calculated using a method involving threshold dependent values). In a pair of studies – one in PET and one in fMRI – Rosen et al. (2011) and McGettigan et al. (2012) took the approach of focusing on the acoustic modulations known to be necessary and sufficient for intelligible percepts – these were variations in formant frequency and amplitude in a form of sinewave speech. In both studies, the authors showed that univariate responses to the two acoustic modulations were bilaterally expressed, when presented alone or in unintelligible combinations. However, the contrast of intelligible with unintelligible stimuli gave a left-dominant response in the univariate analysis, and McGettigan et al. (2012) were further able to demonstrate, using SVM classification, that the left STG and STS contained more information in this classification than the right-hemisphere homologue regions.

Peelle (2012) argues that hemispheric lateralisation of speech processing depends on how speech is defined, making a distinction between “unconnected” (everything up to single words) and “connected” speech (phrases, sentences and narratives). In Peelle’s view, it is the syntactic and semantic aspects of connected speech that engage the left-dominant language

system, but that word comprehension *per se* is a bilaterally mediated process. A number of studies have reported a left-dominant trend in the temporal lobe’s responses to intelligible speech at the level of words and sub-lexical units (e.g. syllables, phonemes; see McGettigan & Scott, 2012 for examples). However, the literature to date is missing an elaboration on the content of representations in the two hemispheres, and how these might differ with regard to “speech-specific” processing. This speaks to the difficulty of disentangling whether preferential responses to speech in temporal regions reflect specialised processing for linguistic inputs, or tuning to the inherent spectro-temporal properties of vocal stimuli (see e.g. Santoro et al., 2014). Indeed, it may prove misguided to assume a privileged status for speech, or for there to be a categorical distinction between speech and other sounds (Iverson, Wagner, & Rosen, 2016; Iverson, Wagner, Pinet, & Rosen, 2011; Rosen & Iverson, 2007).

In sum, the work of the last decade has suggested a clear left-hemisphere dominance for the comprehension of connected speech (i.e. sentences), with additional evidence at lower levels of the linguistic hierarchy. A challenge for future work will be to demonstrate a distinction (or equivalence) in local representations and/or processing between the two hemispheres – this will allow exploration of whether observed asymmetries in the temporal lobes reflect interactions between auditory cortical fields with asymmetric higher-order language-processes (e.g. syntactic processing in left IFG, see Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010).

Contextualising perisylvian responses to intelligible speech within, and alongside, wider neural networks

Early fMRI studies on speech comprehension showed engagement of regions outside of the temporal lobes, including the inferior frontal gyrus (e.g. Binder et al., 2004; Davis & Johnsrude, 2003). In a meta-analysis of the speech comprehension literature, Adank (2012) identified activation of clusters in SMA, precentral gyrus and the left fusiform gyrus by intelligible speech signals. In later work employing multivariate approaches, authors have reported significant classification of intelligible and unintelligible signals in left inferior frontal gyrus and the inferior parietal cortex (Abrams et al., 2013; Evans et al., 2014). Other recent work in which intelligibility and sentence predictability have been co-varied has shown that the inferior frontal and parietal cortices play important roles in semantic facilitation, particularly when speech is moderately degraded (Davis,

Ford, Kherif, & Johnsrude, 2011; Obleser & Kotz, 2010; Obleser, Wise, et al., 2007). A crucial question is whether these regions, particularly those beyond perisylvian cortex, are involved in the core aspects of speech comprehension (i.e. direct mapping between auditory input and intelligible linguistic representations) or if they form more of a “supporting cast”, engaged in downstream computations (e.g. semantic prediction or response selection) under particular tasks or listening contexts. This question was brought to the fore during a resurgence of the Motor Theory of speech perception that was somewhat inspired by attempts to characterise mirror systems in the human brain (see Hickok, 2010; Lotto, Hickok, & Holt, 2009; Pulvermüller & Fadiga, 2010; Scott, McGettigan, & Eisner, 2009). Having observed premotor cortical responses to hearing speech, some speculated that motor representations in the brain play a functional role in speech perception (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). Others suggested that such responses could rather be a result of general behavioural readiness to engage with sounds (e.g. in conversational turn-taking), a consequence of specific task demands such as phonemic segmentation, or a compensation for challenging listening conditions (Scott et al., 2009), and thus that they could reflect a modulatory rather than essential role for motor representations in speech perception (Hickok, Houde, & Rong, 2011).

Several studies employed fMRI to advance the understanding of the role of motor representations, with notable empirical outcomes. Agnew, McGettigan, and Scott (2011) carried out a passive listening study with carefully selected plosive sounds; these were all producible and familiar to the listeners, but were either native or non-native to their spoken language. Here, using a standard univariate subtraction approach, the authors showed a preferential left STS response to native sounds, but no overall difference in the response of motor cortices to the two categories of mouth sound, nor to a baseline condition of signal-correlated noise. Some MVPA studies have also failed to implicate motoric contributions to speech perception (Arsenault & Buchsbaum, 2015, 2016). Others identified a possible role for motoric representations in the processing of degraded speech stimuli, with the implication that articulatory strategies may help listeners to identify speech content when it is unclear or ambiguous (Du et al., 2014; Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012; Lee et al., 2012). Hervais-Adelman et al. (2012) showed increased activation of the left IFG and precentral gyrus when neural responses to degraded but intelligible six-band vocoded speech were compared with clear speech. In an MVPA study, Lee et al. (2012)

examined neural responses to sounds drawn from a continuum from /ba/ to /da/, showing that inferior frontal cortex and pre-SMA contributed to categorising these ambiguous sounds. In agreement, Du et al. (2014) showed highly discriminable neural responses for different spoken syllables in ventral premotor cortex and posterior STG for clear speech, but in the presence of higher levels of competing noise only inferior frontal regions maintained high discriminability. Using DCM, Osnes, Hugdahl, and Specht (2011) found evidence for bi-directional connections between premotor cortex and the STS, but only when listening to ambiguous speech (and not musical chords), suggestive of a possible mechanism by which motoric information might support perception.

A very recent study employing multivariate representational similarity analysis (RSA; for tutorial reviews, see Kriegeskorte et al., 2008 and Nili et al., 2014) offers a new perspective on the nature of the information represented within premotor cortex during speech perception. RSA is a multivariate approach in which dissimilarity distances (typically correlations) are computed between the neural responses associated with each experimental condition. This generates a representational dissimilarity matrix (RDM) that describes the relationship between each condition and every other. The similarity between the RDM and hypothetical models expressing the relation between conditions is then tested to explore different hypotheses. RSA can be thought of in some ways as a more generalised version of MVPA – it can detect binary discriminations between conditions, but in addition, because it uses continuous distances, can detect more nuanced relationships between experimental conditions (Kriegeskorte & Kievit, 2013), resulting in arguably greater reliability than pattern classification approaches (Walther et al., 2016). Other advantages include the ability to test models derived from a wide variety of data sources, and the ability to conduct condition-rich experiments that allow a larger number of hypotheses to be tested (Kriegeskorte et al., 2008). Using RSA, Evans and Davis (2015) showed that premotor cortex contained the most highly abstracted representations of heard syllables (i.e. sensitivity to syllable identity only, whereas the temporal lobes were additionally sensitive to variations in speaker and surface acoustic form). Nonetheless, these authors do not necessarily claim an essential perceptual role for motor representations in the comprehension of speech – rather, they argue for a top-down contribution, which may take a predictive form similar to an efference copy of articulatory information during speech production, and which is more pronounced under challenging listening conditions.

There is a large neuropsychological literature on the importance of the ATL in semantic comprehension, yet functional imaging studies have reported few activations in this region (Visser et al., 2010). Since the establishment of fMRI over PET as the neuroimaging method of choice for studies of healthy brain function, this trend is strongly apparent in studies of speech comprehension. In general, the inferior temporal lobe has received relatively little attention in studies of speech, yet there are clues to its involvement in studies contrasting the perception of auditory speech at varying levels of intelligibility – for example, Adank's (2012) meta-analysis, which implicates a node in left fusiform gyrus, and Halai et al.'s (2015) finding of left inferior temporal and fusiform gyrus activation when employing dual-echo acquisition to reduce signal drop-out artefact. Beyond the issue of whether methodological limitations of fMRI have obscured the true apex of the temporal speech processing hierarchy in the ATL, the question arises as to whether this region reflects a language-specific response, or a domain-general semantic “hub” (Patterson et al., 2007). Very recent analysis of resting state connectivity from the left and right ATLs supports the former claim; this showed that the left ATL engages with ipsilateral perisylvian language regions in frontal and parietal cortex, while the right ATL demonstrates no such connections (Hurley, Bonakdarpour, Wang, & Mesulam, 2015).

In sum, recent work in fMRI has gone beyond a narrow focus upon Broca's and Wernicke's areas to explore the involvement of other structures in supporting speech comprehension. While the premotor cortex may have received somewhat disproportionate attention in the wake of a surging interest in mirror neurons, we describe how the claims from early studies have been addressed and refined through improved experimental designs and the increased sensitivity of multivariate analyses. When considering comprehension, the inferior and most anterior parts of the temporal lobes have been relatively overlooked despite their key associations with semantic processing – improvements to fMRI acquisition protocols, such as dual-echo imaging, should yield improved insights in future work.

Speech comprehension in the reorganising/compensating brain: insights from neuropsychology and healthy ageing populations

Work in neuropsychology has the potential to shine light on our understanding of the neural substrates of speech and language comprehension, for example, through the study of functional reorganisation and responses to therapy after stroke. The degree of preservation of ability, or recovery of behaviour, in aphasic patients

can inform on issues of hemispheric asymmetries in speech comprehension – for example, in the absence of left perisylvian cortex following stroke, can right-hemisphere homologues assume the functions once performed in the left side of the brain? A number of competing views have been presented, suggesting that residual abilities in aphasia might be positively influenced by greater activation of the right hemisphere (Mohr, Difrancesco, Harrington, Evans, & Pulvermüller, 2014) or by the retained function of surviving tissue in the left hemisphere (Crinion & Leff, 2015). However, the role of right-hemisphere structures in this is unclear – a number of studies have found that the degree of right IFG activation is not correlated with behavioural performance in aphasia, and that right-hemisphere activations suggestive of reorganisation may in fact reflect the recruitment of domain-general systems associated with executive control during the performance of difficult tasks (e.g. the right IFG in the cingulo-opercular network, and the right posterior STS/temporoparietal junction in attentional processing; see Geranmayeh, Wise, Mehta, & Leech, 2014). Evidence for the engagement of such non-speech-specific compensatory responses to support speech comprehension has been seen in elevated responses within regions including the anterior cingulate and medial frontal cortex, in both aphasic (Brownsett et al., 2014) and ageing listeners (Erb & Obleser, 2013). With relevance to the role of compensatory right-hemisphere mechanisms, a study of syntactic processing after aphasia found that left-hemisphere stroke patients showed upregulation of activation in right IFG (and middle temporal gyrus; MTG) relative to controls during passive speech comprehension, but it was the amount of activation and tissue preservation in the left IFG that were associated with performance in syntactic processing tasks (Tyler, Wright, Randall, Marslen-Wilson, & Stamatakis, 2010). This supported an earlier study by the same group on syntactic processing in healthy ageing (Tyler, Shafto, et al., 2010).

Thus, the use of clinical and ageing populations has been informative in addressing key empirical issues such as hemispheric asymmetries, namely the capacity for regions in the right hemisphere to perform the same computations as the core left-hemisphere perisylvian language network in speech comprehension.

Future directions

Will fMRI continue to be the imaging technique of choice for localisation of speech networks? In recent years, electrocorticography (iEEG), in which electrical signals are measured directly from the cortical surface, has provided unique insights into the neural basis of speech

perception at a level of temporal and spatial precision that far exceeds fMRI (see Nourski et al., 2013, 2014 for auditory applications, and for speech see Chang et al., 2010; Leonard & Chang, 2014; Mesgarani & Chang, 2012; Mesgarani, Cheung, Johnson, & Chang, 2014; Pasley et al., 2012). These studies have made important contributions to improving our understanding of neural coding within peri-auditory regions, for example, in showing neural tuning to phonetic features rather than to specific phonemes (Mesgarani et al., 2014) and distributed patterns of activity that are strongly associated with representations of speech derived from speech spectrograms (Pasley et al., 2012). Whilst these studies provide a more direct window upon neural processes, we anticipate that the availability of fMRI and the advantages in whole-brain non-invasive acquisition will ensure that it continues to play an important role in research into the near future.

Rather than being superseded by other techniques, we envisage that fMRI will be combined more readily with other neural measures. In the last ten years, studies with magneto/electro-encephalography (M/EEG) have shown the importance of neural oscillations to brain function, providing an alternative and often convergent perspective upon the neural processes involved in speech comprehension. For example, Peelle, Gross, and Davis (2013) showed in MEG that neural oscillations tracking the envelope of speech (between 4 and 7 Hz) were bilateral in response to unintelligible speech, but showed greater phase locking and left lateralisation in the case of intelligible signals (though see Millman, Johnson, and Prendergast (2015) for contradictory evidence). Other studies have shown evidence for hierarchical organisation of oscillatory activity across frequency bands (Lakatos et al., 2005) and the importance of alpha oscillations (8–13 Hz), localised to temporal as well as parietal and prefrontal cortex, in predicting comprehension ratings of degraded speech (Obleser & Weisz, 2012), strongly paralleling findings from fMRI.

In recent years, researchers have attempted to combine electrophysiological and BOLD measures to understand how phenomena described in one modality translate to the other (Becker, Reinacher, Freyer, Villringer, & Ritter, 2011; Hanslmayr et al., 2011). In the future, we suggest that this multimodal approach may be best facilitated by RSA (see Cichy, Pantazis, & Oliva, 2016, for a recent example of this approach in synthesising MEG and fMRI findings). As correlation distances are inherently abstracted from the original signal, different modalities can be related to one another easily without needing to directly understand the mapping between them – the so called “dissimilarity trick” (Kriegeskorte & Kievit, 2013). This allows neural representations across

imaging modalities to be easily compared without requiring data in the two modalities to be collected at the same time, thus offering a flexible means by which to combine electrophysiological responses with fMRI measurements as well as other kinds of physiological measures. For example, one could relate physiological measures of speech production, e.g. sonography or real time MRI, with fMRI measures of speech perception, to gain a more direct understanding of the relation between production and perception (Carey & McGettigan, 2016). In addition, we anticipate that RSA will contribute to advances in the sensitivity with which behavioural data can be related to neuroscientific measures. Indeed, there is a rich history of behavioural research using similarity distances to quantify perceptual relationships among speech sounds (Iverson et al., 2003; Iverson & Kuhl, 1995, 1996; Miller & Nicely, 1955) in ways that more closely capture the multidimensional nature of the speech signal (Kluender & Alexander, 2008; Scott & Evans, 2010). A similarly large literature has applied computational modelling to “lower order” auditory and “higher order” speech comprehension processes (Jepsen, Ewert, & Dau, 2008; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris & McQueen, 2008; Patterson, Allerhand, & Giguere, 1995). RSA allows an intuitive method for linking between these data and neurometric measures (cf. Arsenault and Buchsbaum (2015)). We anticipate that this synthesis might help to better integrate cognitive models and neuroscientific data, for example, in developing models that incorporate neurobiological constraints, and neural accounts that draw more closely from computational models (Fitch, 2014; Gagnepain, Henson, & Davis, 2012). The integration of computational models with fMRI data in particular may help the field to move beyond black-box descriptions of the phenomena of hierarchical processing, toward specifying mechanistic descriptions of the transformations that occur as information proceeds through the dorsal and ventral processing streams.

Whilst RSA may provide further insights into the nature of “local” representational structure, other techniques may be better suited to describing large scale, network-based activity. Existing studies have often used DCM to assess network connectivity. The advantage of this approach is that it allows the researcher to assess how activity in one area causally modulates activity in another, however, DCM requires strong researcher-led assumptions about the architecture of the system under investigation, and analysis can become computationally expensive with increasing numbers of network nodes. An alternative approach is independent component analysis (ICA), which takes advantage of fluctuations in the data to separate the neural signal into

maximally independent spatial maps, or functional networks, in which neural activity is coherently modulated. These components are extracted in a data driven manner, with the timecourse of each component regressed against the experimental design to establish the component's task relatedness (Calhoun, Adali, Pearlson, & Pekar, 2001; see examples from Braga, Wilson, Sharp, Wise, & Leech, 2013; Kamourieh et al., 2015). Whilst ICA is unable to assess causal relations in the functional connectivity between regions, it does have the advantage that it can be conducted on the whole brain without prior assumptions about the underlying structure of functional networks. Using ICA as an initial step to identify the nodes in a functional network, effective connectivity methods such as DCM and Granger causality can be employed as a secondary step to test for causal modulations of activity within the network, gaining maximal benefit from both methods. ICA often identifies additional neural activity not evident in univariate analyses (Geranmayeh, Wise, et al., 2014; Simmonds, Leech, Collins, Redjep, & Wise, 2014). This likely reflects the fact that the same voxel can contribute to multiple overlapping functional networks: task-related activity within these overlapping networks can be modulated in opposite directions, hence differential responses can sometimes be obscured in univariate analyses that effectively average over these independent networks (Xu et al., 2013). By fractionating neural activity into multiple statistically independent components, ICA allows network activity associated with cognitive control and attention networks to be cleanly separated from lower level sensory processing. This may be a useful approach for studying effortful listening, which has previously been shown to engage both domain-general attentional and domain-specific sensory systems (Binder et al., 2004; Evans et al., 2016; Vaden et al., 2013; Wild et al., 2012). Furthermore, the ability to de-mix neural signals into functionally independent networks may be of particular benefit in complex experimental designs in which multiple events occur simultaneously within a trial, for example in "cocktail party listening" or speaking in noise (Braga et al., 2013; Evans et al., 2016; Kamourieh et al., 2015; Meekings et al., 2016). Finally, an additional advantage of ICA is that components from a particular study can be compared to those extracted from large scale studies of resting state networks by correlating the spatial maps (Smith et al., 2009). We anticipate that, in future years, ICA will make interesting contributions to reinterpreting dual stream models of speech perception by redefining the boundaries between linguistic and attentional processes in speech perception (cf. Brownsett et al. (2014)) and establishing a more exact role for the multiple attentional networks engaged

during speech comprehension (Corbetta & Shulman, 2011; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Evans et al., 2016; Petersen & Posner, 2012). Further advances are also likely to come from approaches that constrain the definition of functional networks dependent on the underlying anatomy – studies using diffusion weighted imaging have identified patterns of structural connectivity supportive of prior evidence for functional segregation within regions such as the insula and rostral temporal lobe (Binney, Parker, & Lambon Ralph, 2012; Cloutman, Binney, Drakesmith, Parker, & Lambon Ralph, 2012), and described subdivisions of the dorsal route for language processing (Cloutman, Binney, Morris, Parker, & Lambon Ralph, 2013). RSA may again prove a useful approach to synthesise this kind of structural and functional data (Lima et al., 2015).

Conclusion

We have offered a relatively brief overview of what we consider to be the key developments in the last 10 years of speech comprehension research using functional neuroimaging. A large volume of work from the past decade has provided substantial gains in understanding, notably in elaborating on processing hierarchies in perisylvian cortex, characterising hemispheric asymmetries in the temporal lobes and describing the roles of structures in a wider speech comprehension network. Looking forward, we propose that the examination of representational similarities will produce more comprehensive accounts that bring together data on functional neuroanatomy with neurocomputational primitives (e.g. cortical oscillations), acoustics and behaviour – these approaches will be particularly crucial in delineating pre-lexical stages in comprehension. Alongside advances in describing local representations, techniques such as ICA and DCM will be essential in disentangling the varied and several processes engaged by spoken language processing to identify and characterise the core mechanisms underlying the recognition of intelligible speech signals.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abrams, D. A., Ryali, S., Chen, T., Balaban, E., Levitin, D. J., & Menon, V. (2013). Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. *Cerebral Cortex*, 23, 1703–1714. doi:10.1093/cercor/bhs165

- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42–54. doi:10.1016/j.neuroimage.2012.07.027
- Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience*, 23(12), 4038–4047. doi:10.1162/jocn_a_00106
- Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *Journal of Neuroscience*, 35(2), 634–642. doi:10.1523/JNEUROSCI.2454-14.2015
- Arsenault, J. S., & Buchsbaum, B. R. (2016). No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychonomic Bulletin & Review*, 23, 1231–1240. doi:10.3758/s13423-015-0988-z
- Bandettini, P., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, 25(2), 390–397. doi:10.1002/mrm.1910250220
- Becker, R., Reinacher, M., Freyer, F., Villringer, A., & Ritter, P. (2011). How ongoing neuronal oscillations account for evoked fMRI variability. *Journal of Neuroscience*, 31(30), 11016–11027. doi:10.1523/JNEUROSCI.0210-11.2011
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, 10(4), 417–429. doi:10.1006/nimg.1999.0480
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295–301. doi:10.1038/nn1198
- Binney, R. J., Parker, G. J. M., & Lambon Ralph, M. A. (2012). Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. *Journal of Cognitive Neuroscience*, 24(10), 1998–2014. doi:10.1162/jocn_a_00263
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3), 389–395. doi:10.1038/nn1409
- Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience*, 34(13), 4548–4557. doi:10.1523/JNEUROSCI.4339-13.2014
- Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bihemispheric foundations for human speech comprehension. *Proceedings of the National Academy of Sciences*, 107(40), 17439–17444. doi:10.1073/pnas.1000531107
- Braga, R. M., Wilson, L. R., Sharp, D. J., Wise, R. J. S., & Leech, R. (2013). Separable networks for top-down attention to auditory non-spatial and visuospatial modalities. *NeuroImage*, 74, 77–86. doi:10.1016/j.neuroimage.2013.02.023
- Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain*, 137(1), 242–254. doi:10.1093/brain/awt289
- Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140–151. doi:10.1002/hbm.1048
- Caplan, D. (2009). Experimental design and interpretation of functional neuroimaging studies of cognitive processes. *Human Brain Mapping*, 30(1), 59–77. doi:10.1002/hbm.20489
- Carey, D., & McGettigan, C. (2016). Magnetic resonance imaging of the brain and vocal tract: applications to the study of speech production and language learning. *Neuropsychologia*, 1–11. doi:10.1016/j.neuropsychologia.2016.06.003
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428–1432. doi:10.1038/nn.2641
- Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex*, 26(8), 3563–3579. doi:10.1093/cercor/bhw135
- Cloutman, L. L., Binney, R. J., Drakesmith, M., Parker, G. J. M., & Lambon Ralph, M. A. (2012). The variation of function across the human insula mirrors its patterns of structural connectivity: Evidence from in vivo probabilistic tractography. *NeuroImage*, 59(4), 3514–3521. doi:10.1016/j.neuroimage.2011.11.016
- Cloutman, L. L., Binney, R. J., Morris, D. M., Parker, G. J. M., & Lambon Ralph, M. A. (2013). Using in vivo probabilistic tractography to reveal two segregated dorsal “language-cognitive” pathways in the human brain. *Brain and Language*, 127(2), 230–240. doi:10.1016/j.bandl.2013.06.005
- Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience*, 34(1), 569–599. doi:10.1146/annurev-neuro-061010-113731
- Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *Journal of Neuroscience*, 34(1), 332–338. doi:10.1523/JNEUROSCI.1302-13.2014
- Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *Journal of Neuroscience*, 35(45), 15015–15025. doi:10.1523/JNEUROSCI.0977-15.2015
- Crinion, J. T., & Leff, A. P. (2015). Using functional imaging to understand therapeutic effects in poststroke aphasia. *Current Opinion in Neurology*, 28(4), 330–337. doi:10.1097/WCO.0000000000000217
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8, 109–114. doi:10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.3.CO;2-N
- Davis, M., Ford, M., Kherif, F., & Johnsrude, I. (2011). Does semantic context benefit speech understanding through “top down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914–3932. doi:10.1162/jocn_a_00084
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423–3431.
- De Martino, F., Moerel, M., Ugurbil, K., Formisano, E., & Yacoub, E. (2015). Less noise, more activation: Multiband acquisition schemes for auditory functional MRI. *Magnetic Resonance in Medicine*, 74(2), 462–467. doi:10.1002/mrm.25408

- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., ... Tyler, L. K. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage*, 11(6), 589–600. doi:10.1006/nimg.2000.0595
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12(3), 99–105. doi:10.1016/j.tics.2008.01.001
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences*, 111(19), 7126–7131. doi:10.1073/pnas.1318738111
- Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, 7(2), 89–97.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, 30(21), 7179–7186. doi:10.1523/JNEUROSCI.4040-09.2010
- Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, 33(26), 10688–10697. doi:10.1523/JNEUROSCI.4596-12.2013
- Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, 7, 116. doi:10.3389/fnsys.2013.00116
- Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex*, 25, 4772–4788. doi:10.1093/cercor/bhv136
- Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., ... Scott, S. K. (2014). The pathways for intelligible speech: Multivariate and univariate perspectives. *Cerebral Cortex*, 24, 2350–2361. doi:10.1093/cercor/bht083
- Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the cocktail party started: Masking effects in speech perception. *Journal of Cognitive Neuroscience*, 28(3), 483–500. doi:10.1162/jocn_a_00913
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, 11(3), 329–364. doi:10.1016/j.plrev.2014.04.005
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? brain-based decoding of human voice and speech. *Science*, 322, 970–973. doi:10.1126/science.1164318
- Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4), 859–869.
- Foster, J. R., Hall, D., Summerfield, A. Q., Palmer, A. R., & Bowtell, R. W. (2000). Sound-level measurements and calculations of safe noise dosage during EPI at 3T. *Journal of Magnetic Resonance Imaging*, 12(1), 157–163. doi:10.1002/1522-2586(200007)12:1<157::aid-jmri17>3.0.co;2-m
- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, 31(3), 448–457. doi:10.1002/hbm.20878
- Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, 4(2), 97–104. doi:10.1006/nimg.1996.0033
- Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, 10(5), 607–619. doi:10.1006/nimg.1999.0498
- Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, 23(2), 172–178. doi:10.1016/j.conb.2012.11.010
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, 22(7), 615–621. doi:10.1016/j.cub.2012.02.015
- Ge, J., Peng, G., Lyu, B., Wang, Y., Zhuo, Y., Niu, Z., ... Gao, J.-H. (2015). Cross-language differences in the brain network subserving intelligible speech. *Proceedings of the National Academy of Sciences*, 112(10), 2972–2977. doi:10.1073/pnas.1416000112
- Geranmayeh, F., Brownsett, S. L. E., & Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: What is driving recovery? *Brain*, 137(Pt 10), 2632–2648. doi:10.1093/brain/awu163
- Geranmayeh, F., Wise, R. J. S., Mehta, A., & Leech, R. (2014). Overlapping networks engaged during spoken language production and its cognitive control. *Journal of Neuroscience*, 34(26), 8728–8740. doi:10.1523/JNEUROSCI.0428-14.2014
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. doi:10.1038/nn.3063
- Guediche, S., Holt, L. L., Laurent, P., Lim, S.-J., & Fiez, J. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, 25(7), 1867–1877. doi:10.1093/cercor/bht428
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research*, 271, 133–146. doi:10.1016/j.heares.2010.01.011
- Halai, A. D., Welbourne, S. R., Embleton, K., & Parkes, L. M. (2014). A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Human Brain Mapping*, 35(8), 4118–4128. doi:10.1002/hbm.22463
- Halai, A., Parkes, L. M., & Welbourne, S. (2015). Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *NeuroImage*, 122, 214–221. doi:10.1016/j.neuroimage.2015.05.067
- Hall, D. A., Chambers, J., Akeroyd, M. A., Foster, J. R., Coxon, R., & Palmer, A. R. (2009). Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *Journal of the Acoustical Society of America*, 125(1), 347–359. doi:10.1121/1.3021437
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., ... Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3), 213–223.
- Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bauml, K.-H. T. (2011). The relationship between brain oscillations and BOLD signal during memory formation: A combined EEG-fMRI study. *Journal of Neuroscience*, 31(44), 15674–15680. doi:10.1523/JNEUROSCI.3140-11.2011

- Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523–534. doi:10.1038/nrn1931
- Herrmann, B., Obleser, J., Kalberlah, C., Haynes, J. D., & Friederici, A. D. (2012). Dissociable neural imprints of perception and grammar in auditory functional imaging. *Human Brain Mapping*, 33(3), 584–595. doi:10.1002/hbm.21235
- Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, 27(7–8), 1145–1166. doi:10.1080/01690965.2012.662280
- Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, 25(6), 749–776. doi:10.1080/01690961003595572
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3), 407–422. doi:10.1016/j.neuron.2011.01.019
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. doi:10.1038/nrn2113
- Hurley, R. S., Bonakdarpour, B., Wang, X., & Mesulam, M.-M. (2015). Asymmetric connectivity between the anterior temporal lobe and the language network. *Journal of Cognitive Neuroscience*, 27(3), 464–473. doi:10.1162/jocn_a_00722
- Iverson, P., & Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1), 553–562. doi:10.1121/1.412280
- Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of r and l. *Journal of the Acoustical Society of America*, 99(2), 1130–1140.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47–B57. doi:10.1016/S0010-0277(02)00198-1
- Iverson, P., Wagner, A., Pinet, M., & Rosen, S. (2011). Cross-language specialization in phonetic processing: English and Hindi perception of /w/-/v/ speech and nonspeech. *Journal of the Acoustical Society of America*, 130(5), EL297–EL303. doi:10.1121/1.3632048
- Iverson, P., Wagner, A., & Rosen, S. (2016). Effects of language experience on pre-categorical perception: Distinguishing general from specialized processes in speech perception. *The Journal of the Acoustical Society of America*, 139(4), 1799–1809. doi:10.1121/1.4944755
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 23(29), 9541–9546.
- Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, 124(1), 422–438. doi:10.1121/1.2924135
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, 97(22), 11793–11799. doi:10.1073/pnas.97.22.11793
- Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, 9(2), 164–170.
- Kamourieh, S., Braga, R. M., Leech, R., Newbould, R. D., Malhotra, P., & Wise, R. J. S. (2015). Neural systems involved when attending to a speaker. *Cerebral Cortex*, 25(11), 4284–4298. doi:10.1093/cercor/bhu325
- Kaplan, J. T., Man, K., & Greening, S. G. (2015). Multivariate cross-classification: Applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, 9(March), 151. doi:10.3389/fnhum.2015.00151
- Kilian-Hutten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, 31(5), 1715–1720. doi:10.1523/JNEUROSCI.4572-10.2011
- Kluender, K. R., & Alexander, J. M. (2008). Perception of speech sounds. *The Senses: A Comprehensive Reference*, 3, 829–860. doi:10.1016/B978-012370880-9.00067-0
- Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4), 649–662. doi:10.1016/j.neuroimage.2007.02.022
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10), 3863–3868. doi:10.1073/pnas.0600244103
- Kriegeskorte, N., & Kievit, R. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412. doi:10.1016/j.tics.2013.06.007
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2(November), 4–10. doi:10.3389/neuro.06.004.2008
- Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., ... Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12), 5675–5679. doi:10.1073/pnas.89.12.5675
- Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: An fMRI study. *Journal of Cognitive Neuroscience*, 26(8), 1748–1763. doi:10.1162/jocn_a_00583
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., Schroeder, C. E., ... Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, 94, 1904–1911. doi:10.1152/jn.00263.2005
- Langers, D. R. M., Van Dijk, P., & Backes, W. H. (2005). Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53(1), 49–60. doi:10.1002/mrm.20315
- Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience*, 32(11), 3942–3948. doi:10.1523/JNEUROSCI.3814-11.2012

- Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2008). The cortical dynamics of intelligible speech. *Journal of Neuroscience*, 28(49), 13209–13215. doi:10.1523/JNEUROSCI.2903-08.2008
- Leonard, M. K., & Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends in Cognitive Sciences*, 18(9), 472–479. doi:10.1016/j.tics.2014.05.001
- Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., ... Scott, S. K. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex*, 25(11), 4638–4650. doi:10.1093/cercor/bhv134
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, a. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157. doi:10.1038/35084005
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, 66(1), 735–769. doi:10.1146/annurev.physiol.66.082602.092845
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110–114. doi:10.1016/j.tics.2008.11.008
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. doi:10.1016/0010-0277(87)90005-9
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. doi:10.1016/0010-0285(86)90015-0
- McGettigan, C., Evans, S., Rosen, S., Agnew, Z. A., Shah, P., & Scott, S. K. (2012). An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience*, 24(3), 636–652. doi:10.1162/jocn_a_00161
- McGettigan, C., & Scott, S. (2012). Cortical asymmetries in speech perception: What's wrong, what's right and what's left? *Trends in Cognitive Sciences*, 16(5), 269–276. doi:10.1016/j.tics.2012.04.006
- Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S. K. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *The Journal of the Acoustical Society of America*, 140(1), 8–19. doi:10.1121/1.4948587
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233–236. doi:10.1038/nature11020
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006–1010. doi:10.1126/science.1245994
- Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352. doi:10.1121/1.1907526
- Millman, R. E., Johnson, S. R., & Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *Journal of Cognitive Neuroscience*, 27(3), 533–545. doi:10.1162/jocn_a_00719
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, 53(1), 103–118. doi:10.1016/j.neuroimage.2010.05.051
- Mohr, B., Difrancesco, S., Harrington, K., Evans, S., & Pulvermüller, F. (2014). Changes of right-hemispheric activation after constraint-induced, intensive language action therapy in chronic aphasia: fMRI evidence from auditory semantic processing. *Frontiers in Human Neuroscience*, 8 (November), 919. doi:10.3389/fnhum.2014.00919
- Mourao-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage*, 28(4), 980–995. doi:10.1016/j.neuroimage.2005.06.070
- Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, 103, 130–138. doi:10.1016/j.neuroimage.2014.09.026
- Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, 4(1), 101–109. doi:10.1093/scan/nsn044
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12), 1362–1368.
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, 10(4), e1003553. doi:10.1371/journal.pcbi.1003553
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430. doi:10.1016/j.tics.2006.07.005
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115 (2), 357–395. doi:10.1037/0033-295X.115.2.357
- Nourski, K. V., Brugge, J. F., Reale, R., Kovach, C. K., Oya, H., Kawasaki, H., ... Howard, M. (2013). Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: An intracranial electrophysiology study. *Journal of Neurophysiology*, 109(319), 1283–1295. doi:10.1152/jn.00718.2012
- Nourski, K. V., Steinschneider, M., Oya, H., Kawasaki, H., Jones, R. D., & Howard, M. (2014). Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. *Cerebral Cortex*, 24(2), 340–352. doi:10.1093/cercor/bhs314
- Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., ... Rauschecker, J. P. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*, 27(7), 562–571. doi:10.1002/hbm.20201
- Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1), 14–19. doi:10.1016/j.tics.2008.09.005
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, 28(32), 8116–8123. doi:10.1523/JNEUROSCI.1290-08.2008
- Obleser, J., & Kotz, S. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3), 633–640. doi:10.1093/cercor/bhp128
- Obleser, J., Leaver, A. M., Vanmeter, J., & Rauschecker, J. P. (2010). Segregation of vowels and consonants in human

- auditory cortex: Evidence for distributed hierarchical organization. *Frontiers in Psychology*, 1, 232. doi:10.3389/fpsyg.2010.00232
- Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex*, 22, 2466–2477. doi:10.1093/cercor/bhr325
- Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, 27(9), 2283–2289. doi:10.1523/JNEUROSCI.4663-06.2007
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17, 2251–2257. doi:10.1093/cercor/bhl133
- Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M., & Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophysical Journal*, 64(3), 803–812. doi:10.1016/S0006-3495(93)81441-3
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., ... Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, 20, 2486–2495. doi:10.1093/cercor/bhp318
- Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage*, 54(3), 2437–2445. doi:10.1016/j.neuroimage.2010.09.078
- O'Toole, A. J., Jiang, F., Abdi, H., Penard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience*, 19(11), 1735–1752. doi:10.1162/jocn.2007.19.11.1735
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S., Crone, N. E., ... Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1), e1001251. doi:10.1371/journal.pbio.1001251
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987. doi:10.1038/nrn2277
- Patterson, R., Allerhand, M., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4), 1890–1894. doi:10.1121/1.414456
- Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: A hierarchical perspective. *Frontiers in Human Neuroscience*, 6(November), 1–4. doi:10.3389/fnhum.2012.00309
- Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, 8(JUL), 1–13. doi:10.3389/fnins.2014.00253
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, 52(4), 1410–1419. doi:10.1016/j.neuroimage.2010.05.015
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. doi:10.1093/cercor/bhs118
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, 45(1), S199–S209. doi:10.1016/j.neuroimage.2008.11.007
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35(1), 73–89. doi:10.1146/annurev-neuro-062111-150525
- Pfeuffer, J., van de Moortele, P.-F., Yacoub, E., Shmuel, A., Adriany, G., Andersen, P., ... Hu, X. (2002). Zoomed functional imaging in the human brain at 7 Tesla with simultaneous high spatial and high temporal resolution. *NeuroImage*, 17(1), 272–286. doi:10.1006/nimg.2002.1103
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1), 245–255. doi:10.1016/S0167-6393(02)00107-3
- Poser, B., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, 55(6), 1227–1235. doi:10.1002/mrm.20900
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847. doi:10.1016/j.neuroimage.2012.04.062
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–360. doi:10.1038/nrn2811
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865–7870. doi:10.1073/pnas.0509989103
- Raizada, R. D. S., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4), 726–740. doi:10.1016/j.neuron.2007.11.001
- Raizada, R. D. S., Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex*, 20(1), 1–12. doi:10.1093/cercor/bhp076
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. doi:10.1038/nn.2331
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22), 11800–11806. doi:10.1073/pnas.97.22.11800
- Rodd, J., Davis, M., & Johnsrude, I. (2005). The neural mechanisms of speech: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261–1269. doi:10.1093/cercor/bhi009
- Rosen, S., & Iverson, P. (2007). Constructing adequate non-speech analogues: What is special about speech anyway? *Developmental Science*, 10(2), 165–168. doi:10.1111/j.1467-7687.2007.00550.x
- Rosen, S., Wise, R. J. S., Chadha, S., Conway, E. J., & Scott, S. K. (2011). Hemispheric asymmetries in speech perception:

- Sense, nonsense and modulations. *PLoS ONE*, 6(9), e24672. doi:10.1371/journal.pone.0024672
- Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Computational Biology*, 10(1), e1003412. doi:10.1371/journal.pcbi.1003412
- Schmitter, S., Diesch, E., Amann, M., Kroll, a., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory fMRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(5), 317–325. doi:10.1007/s10334-008-0132-4
- Schonwiesner, M., Rubsamen, R., & von Cramon, D. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, 22(6), 1521–1528. doi:10.1196/annals.1360.051
- Schwarzbauer, C., Davis, M. H., Rodd, J. M., & Johnsrude, I. (2006). Interleaved Silent Steady State (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, 29(3), 774–782. doi:10.1016/j.neuroimage.2005.08.025
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400–2406. doi:10.1093/brain/123.12.2400
- Scott, S. K., & Evans, S. (2010). Categorizing speech. *Nature Neuroscience*, 13(11), 1304–1306. doi:10.1038/nn1110-1304
- Scott, S., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295–302. doi:10.1038/nrn2603
- Simmonds, A. J., Leech, R., Collins, C., Redjep, O., & Wise, R. J. S. (2014). Sensory-motor integration during speech production localizes to both left and right plana temporale. *Journal of Neuroscience*, 34(39), 12963–12972. doi:10.1523/JNEUROSCI.0336-14.2014
- Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31), 13040–13045. doi:10.1073/pnas.0905267106
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19(6), 498–502. doi:10.1016/j.cub.2009.01.066
- Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, 49(4), 3099–3109. doi:10.1016/j.neuroimage.2009.11.015
- Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, 77, 157–165. doi:10.1016/j.neuroimage.2013.03.039
- Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: The modulation of the fronto-temporal language system in the context of age-related atrophy. *Cerebral Cortex*, 20(2), 352–364. doi:10.1093/cercor/bhp105
- Tyler, L. K., Wright, P., Randall, B., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Reorganization of syntactic processing following left-hemisphere brain damage: Does right-hemisphere activity preserve function? *Brain*, 133(11), 3396–3408. doi:10.1093/brain/awq262
- Upadhyay, J., Silver, A., Knaus, T. A., Lindgren, K. A., Ducros, M., Kim, D. S., & Tager-Flusberg, H. (2008). Effective and structural connectivity in the human auditory cortex. *Journal of Neuroscience*, 28(13), 3341–3349. doi:10.1523/JNEUROSCI.4434-07.2008
- Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal of Neuroscience*, 33(48), 18979–18986. doi:10.1523/JNEUROSCI.1417-13.2013
- Visser, M., Jefferies, E., & Lambon Ralph, M. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083–1094. doi:10.1162/jocn.2009.21309
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 137, 188–200. doi:10.1016/j.neuroimage.2015.12.012
- Watkins, K. E., Paus, T., Lerch, J. P., Zijdenbos, a., Collins, D. L., Neelin, P., ... Evans, A. C. (2001). Structural asymmetries in the human brain: A voxel-based statistical analysis of 142 MRI scans. *Cerebral Cortex*, 11(9), 868–877. doi:10.1093/cercor/11.9.868
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32(40), 14010–14021. doi:10.1523/JNEUROSCI.1528-12.2012
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702. doi:10.1038/nn1263
- Xu, J., Zhang, S., Calhoun, V. D., Monterosso, J., Li, C.-S. R., Worhunsky, P. D., ... Potenza, M. N. (2013). Task-related concurrent but opposite modulations of overlapping functional networks as revealed by spatial ICA. *NeuroImage*, 79, 62–71. doi:10.1016/j.neuroimage.2013.04.038
- Yacoub, E., Duong, T. Q., Van De Moortele, P.-F., Lindquist, M., Adriany, G., Kim, S.-G., ... Hu, X. (2003). Spin-echo fMRI in humans using high spatial resolutions and high magnetic fields. *Magnetic Resonance in Medicine*, 49, 655–664. doi:10.1002/mrm.10433
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946–953.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1087–1104. doi:10.1098/rstb.2007.2161