

# How to study spoken language understanding: a survey of neuroscientific methods

Jennifer M. Rodd<sup>1</sup> and Matthew, H. Davis<sup>2</sup>

1. Department of Experimental Psychology, University College London, UK, WC1H 0AP. Email: [j.rodd@ucl.ac.uk](mailto:j.rodd@ucl.ac.uk)

2. MRC Cognition and Brain Sciences Unit, Cambridge, UK.

## Abstract

The past 20 years have seen a methodological revolution in spoken language research. A diverse range of neuroscientific techniques are now available that allow researchers to observe the brain's responses to different types of speech stimuli in both healthy and impaired listeners, and also to observe how individuals' abilities to process speech change as a consequence of disrupting processing in specific brain regions. This special issue provides a tutorial review of the most important of these methods to guide researchers to make informed choices about which methods are best suited to addressing specific questions concerning the neuro-computational foundations of spoken language understanding. This introductory review provides (i) an historical overview of the experimental study of spoken language understanding, (ii) a summary of the key method currently being used by cognitive neuroscientists in this field, and (iii) thoughts on the likely future developments of these methods.

## 1. How to study spoken language understanding

The ability to communicate effectively with other people using spoken language is a fundamental human ability that has profound, long-term consequences for an individual's success in life, both in terms of measures of academic attainment and occupational status (Johnson, Beitchman, & Brownlie, 2010). For over 100 years scientists have attempted to understand the specific nature of the mechanisms that support successful spoken language comprehension from both cognitive and neural perspectives. This increased understanding of the neurobiology of spoken language comprehension provides an essential foundation for the development of successful interventions for children with developmental language disorders (Krishnan, Watkins, & Bishop, 2016) and for individuals who have acquired speech processing deficits as a consequence of stroke and other brain injuries (Saur & Hartwigsen, 2012). Understanding the neurocognitive mechanisms that support speech comprehension is also essential for fully understanding other forms of communication, such as reading and sign-language.

From a cognitive perspective, the endeavour to understand how spoken words were recognised and understood was revolutionised in the early 1970s when researchers began to develop a set of experimental tools that provided a window on how one specific word (e.g., CAPTAIN) might be recognised from the cohort of similar sounding words (e.g., CAPTIVE). In highly influential set of studies William Marslen-Wilson used a word shadowing paradigm in

which listeners were required to repeat back spoken sentences as rapidly as possible (Marslen-Wilson, 1973; Marslen Wilson, 1975). These experiments showed that some listeners are able to repeat back continuous speech at delays of only 250 msec. Crucially, even at these short latencies, any errors that listeners made were syntactically and semantically constrained, such as adding in appropriate (but missing) function words. Listeners also corrected the pronunciation of mispronounced words. These results show that participants were not simply parroting back the sounds that they were hearing but were recognizing words, retrieving their meanings and then repeating them back within approximately 250 msec. These findings demonstrated for the first time the remarkable speed of speech comprehension, putting pressure on psycholinguists to find research tools capable of revealing the underlying mechanisms that comprehend incoming speech so quickly and efficiently.

By the mid-1990s researchers had responded to this challenge by producing a range of experimental tools suitable for addressing a range of research questions about how speech is understood. In 1996 a group of influential researchers joined forces to publish a special issue of this journal that catalogued the different methods being used to study spoken word recognition (Grosjean & Frauenfelder, 1996). Each chapter of this special issue focused on a single experimental method, many of which are still in use today. In some of these methods, participants report (verbally) the content of the speech that they have heard and are then scored on the accuracy of these reports (e.g., gating: Grosjean, 1996; word identification in noise: Pisoni, 1996). A second

common approach involves measuring the speed and accuracy of participants' forced-choice button press responses to speech stimuli (e.g., auditory lexical decision: Goldinger, 1996; word monitoring: Kilborn & Moss, 1996). Finally, in priming tasks, the incidental impact of previous presentations of spoken materials is inferred from facilitation of responses to subsequent written or spoken materials (e.g. form priming: Zwitserlood, 1996; cross-modal semantic priming: Tabossi, 1996). Critically, many of these methods were developed to reveal not only the *output* of the speech comprehension system (i.e. *which* word the listener perceived), but also the time-course with which this information was accessed. These methods aimed to provide a window onto speech comprehension – researchers were able to 'sneak a look' at the usually invisible process by which a listener transforms the physical sound stimulus into an internal, meaningful representation to which only the listener would usually have access.

The 20 years since the publication of this special issue have seen a second methodological revolution in spoken language research: there has been a rapid expansion in the methods available to study the *neural* basis of speech processing. Historically, the primary source of information about how different brain regions contribute to specific aspects of speech was patients with speech processing difficulties, but more recently a range of technological developments have provided researchers with a range of different tools for observing responses of the brain to different types of speech stimuli in both healthy and impaired listeners as well as observing how individuals' ability to process speech can change as a consequence of temporarily disrupting processing in specific brain regions (Passingham & Rowe, 2015).

These diverse approaches to studying how the brain processes speech can provide various different kinds of information that constrains our theories of how spoken language is processed. First, they can be used to answer what can be thought of as strictly neurobiological questions – questions about *where* in the brain specific types of representations or processes might be instantiated. Second, some current neuroimaging techniques provide a dependent measure that can be used to answer strictly cognitive questions – just as differences in response times between conditions can provide insights into the cognitive mechanisms by which different types of stimuli are processed, so can differences in the magnitude or timing of the neural response (see Henson, 2005). Indeed in some cases, neuroscientific dependent measures have advantages compared with more traditional behavioural measures: just as studies of eye-movements allow researchers interested in reading to observe participants' responses in a relatively naturalistic task-free environment, neuroimaging methods such as fMRI, EEG and MEG can be used to directly observe the

changes in neural activity that occur during comprehension of different types of speech without necessarily 'contaminating' these observations by requiring participants to make additional explicit, meta-linguistic decisions about the speech that they heard (e.g., lexical decision, semantic categorization). Similarly, neuroimaging can be used to study spoken language comprehension under circumstances that are difficult (or impossible) using methods that require a behavioural response, such as when participants are sedated (Davis et al., 2007) or for brain-injured participants who are unable to make overt responses to speech (Coleman et al., 2007; Coleman et al., 2009)<sup>1</sup>. Finally, in addition to answering strictly neural or cognitive questions, by combining behavioural and neural measures, the diverse set of neuroscientific methods that are now available can (potentially) allow for far richer mechanistic theories that explain the underlying cognitive processes as arising from specific neural computations that can be shown to operate in specific brain areas.

It is this last application of neuroscience methods that provides a critical motivation for the present special issue. The set of neural methods that are described in this special issue have now developed to the point at which they are increasingly able to constrain and inform theorising so as to pave the way for unified cognitive and neuroscientific theories of language comprehension. The aim of this special issue is to provide a tutorial review of the most important of these methods. Our focus here is not on theory, but on the methodological issues that arise for researchers interested in studying speech comprehension. We hope that this special issue will guide researchers to make informed choices about which methods are best suited to addressing specific questions concerning the neuro-computational foundations of spoken language understanding.

## 2. Experimental challenges in studying speech comprehension

Speech has four characteristics that present specific challenges to researchers that are not universally present in other areas of experimental psychology or cognitive neuroscience. Firstly, speech is an auditory stimulus. This obvious, but intrinsic, characteristic of speech presents a fundamental challenge when using cognitive neuroscience methods that are themselves inherently noisy. For example, MRI scanners produce continuous noise of more than 90 dB SPL during image acquisition (Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010), and the discharge of a TMS coil can be similarly loud (Dhamne et al., 2014). In the behavioural literature on speech understanding researchers typically work hard to achieve high-fidelity presentation of clear speech or in some cases use the presence of background noise to deliberately perturb spoken language understanding (Pisoni, 1996). Speech

presented in conjunction with noisy neuroscientific methods necessarily leads to challenges or compromises in experimental design – researchers should either acknowledge that they are studying speech comprehension in the presence of significant background noise, or use sparse or offline methods in which speech presentation is timed to avoid noisy periods of data collection or brain stimulation (Perrachione & Ghosh, 2013; Devlin & Watkins, 2007; Schwarzbauer, Davis, Rodd, & Johnsrude, 2006; Hall et al., 1999; Peelle, 2014). In many cases, however, the additional methodological issues that arise for auditory, but not visual stimuli, have resulted in researchers taking the easier but rather limiting approach of studying language comprehension using written rather than spoken language. This imbalance is most apparent for higher level aspects of speech/language comprehension such as grammatical processing, where the majority of research has been carried out with visually presented words (see Rodd, Vitello, Woollams, & Adank, 2015).

A second property of speech that constrains our experimental designs is that it is inevitably a continuous signal that is distributed in time. While a written word can be presented instantaneously to the participants who can process its entire visual form simultaneously, spoken words unfold over time with their initial sounds being heard before the later parts of the word. The majority of the behavioural methods used to study speech deal with this issue by forcing a discrete response at a specific time point, and thereby obtaining a single snapshot of processing at that precise point in time. An alternative approach, that can potentially provide far richer insights into the time-course of speech processing, is to use a method that provides a continuous outcome measure of processing. One relatively rare example of this from the cognitive literature is the visual world method in which listeners' eye-movements are measured while they hear a sentence that refers to objects in the visual scene (Tanenhaus & Spivey-Knowlton, 1996). This method provides a continuous measure of the degree to which different perceptual hypotheses are activated, with the constraint that only a few perceptual interpretations can be assessed in a single trial and that all the words used should refer to picturable objects. In contrast, with neuroscientific measures it is relatively common to acquire a continuous measure of the brain's response (e.g., fMRI: Evans & McGettigan, 2017; MEG/EEG: Wöstmann., Fiedler, & Obleser, 2016). However, the temporal nature of speech adds considerable complexity to such experiments. The brain's responses to visually presented words can be measured from the onset of visual presentation so that researchers can be certain that the observed time-course of neural responses reflects the relatively orderly sequence of perceptual and cognitive processes involved in word recognition. However, interpretation of speech-

evoked neural responses is much more challenging. The observed time-course of neural responses will be driven both by the time taken for perceptual/cognitive processes involved in spoken word recognition, but also by the time-course of the speech signal itself. For example, a neural response observed around the offset of the word could reflect a relatively slow response to the initial speech sounds, a more rapid response to sounds heard immediately prior to the offset of the word, or even a preparatory response to subsequently presented words. Although carefully constructed experiments can allow experimenters to separate out the responses that are being driven by different components of the unfolding speech stimulus (e.g., Zwitserlood & Schriefers, 1995; O'Rourke & Holcomb, 2002; Lerner, Honey, Katkov, & Hasson, 2014; Vagharchakian, Dehaene-Lambertz, Pallier, & Dehaene, 2012; Ahissar et al., 2001), this additional complexity continues to present challenges to speech researchers who are aiming to characterise the temporal profile of the different component stages of speech perception/comprehension.

A third, and closely related, property of speech that can be challenging for researchers is the considerable variation in the duration and timing of individual speech tokens: not only do spoken words (in general) unfold over time, but different words unfold with highly variable, idiosyncratic timing profiles. This unavoidable variation across stimulus items can be highly problematic for speech researchers. Consider again the researcher setting up an experiment using visually presented single words. This researcher would be able to minimise the nuisance within-condition variance by selecting words with the same number of letters and then presenting these words on screen for an identical amount of time. In contrast, for the analogous auditory experiment where the researcher was using recorded tokens of speech from a human speaker, even if these words were carefully controlled for the number of constituent speech segments there would be considerable natural variation in the duration of the individual speech tokens. Even if the researcher elected to edit these speech stimuli such that they had a consistent overall duration, each individual word would have a unique internal time-course in terms of the rate at which the constituent sounds occurred. Perhaps most significantly, there will be natural variation in the point at which the listener has heard enough to be able to uniquely identify that word from its cohort of similar sounding competitors (e.g. distinguishing "captain" from "captive"; Marslen-Wilson, 1984; Davis & Rodd, 2011). Similarly, while it is possible to construct auditory sentence materials that are relatively well controlled in terms of their total duration, there will inevitably be considerable natural variability in terms of the exact timing of the lexical (and sub-lexical) events within the sentence. Although this variation in the time-course of events within speech stimuli raises issues for all experimental

studies of speech, including both behavioural and neuroscientific methods, it is particularly problematic for methods that depend on neural responses being time-locked to a specific event and then averaged across trials; researchers need to commit to a specific point in time at which an equivalent neural response is measured (in practice, often the uniqueness or divergence point of the speech stimulus is used; e.g., O'Rourke & Holcomb, 2002; Gagnepain, Henson, & Davis, 2012; MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012; Kocagoncu, Clarke, Devereux, & Tyler, 2017). The inevitable variability in the timing of the brain's response that is driven by differences in the rate of neural processing for different stimulus items or different participants will significantly reduce the signal-to-noise ratio for such studies compared to an analogous study of visually presented words.

A final set of methodological issues arise because, unlike text, natural speech always comes from a single specific speaker. In reading studies printed words are usually presented in highly familiar standard fonts. In contrast for speech experiments, the speaker's voice is usually unfamiliar to participants. It is well known that there are significant differences in how listeners process speech from familiar and unfamiliar speakers, and importantly that they can adapt relatively rapidly within the course of an experiment to new speakers, with changes in the accuracy of speech processing as the listener becomes more familiar with the particular speaker, especially for speech presented within background noise (e.g., Mullennix, Pisoni, & Martin, 1989; Nygaard & Pisoni, 1998). It is therefore possible that in some experiments, participants' performance may change during the experiment in ways that would not occur in an analogous reading experiment. While perceptual variation might only add variability to the data, it remains unclear whether this issue might potentially produce consistent confounds, such that (for example) qualitatively different results might be observed in long vs. short experiments. In addition, while most researchers avoid using speech that their participants consider to be strongly accented, spoken language is always produced with a specific accent that can contain significant clues about the speaker's gender, age, social class or education level. This information can directly influence listeners' processing of speech within experimental contexts in ways that are mostly absent for text (Van Berkum, Van Den Brink, Tesink, Kos, & Hagoort, 2008; Martin, Garcia, Potter, Melinger, & Costa, 2016). Speech researchers should therefore keep in mind that, even in relatively low-level speech perception experiments, participants interpret the stimuli within a broader linguistic context in which the speaker is viewed as a social agent (Hay & Drager, 2010). A final issue that arises due to speaker differences is that even for pairs of studies that are being conducted in the same language, it is often inappropriate to use the same speech tokens in

experiments conducted in different geographical locations where different accents will be the norm. (Note that a similar issue arises to a lesser extent, for studies of reading where different dialects may differ in vocabulary and spelling). This aspect of speech can constrain reproducibility across labs as stimuli must necessarily be rerecorded with a locally appropriate accent.

In summary, researchers interested in understanding the neuro-cognitive basis of speech processing face significant methodological challenges that are a consequence of the nature of speech itself. These factors must be kept in mind both when choosing an appropriate experimental technique, and when designing specific experiments.

### **3. Overview of cognitive neuroscience methods for studying spoken language understanding**

Investigations of the brain systems supporting spoken language understanding can adopt one of two broad approaches illustrated in Figure 1; (1) brain imaging and (2) neuropsychology/brain stimulation. In brain imaging experiments the researcher varies (usually as the independent variable) either the speech stimuli heard by participants, or the behavioural response that is required in response to these stimuli, and observes the consequent changes in brain activity. For experiments on speech comprehension, common experimental manipulations might be to compare speech stimuli that are comprehended or not comprehended due to auditory degradation, or that vary in the ease of comprehension due to the presence/absence of lexical or semantic anomaly or ambiguity (e.g., Scott, Blank, Rosen, & Wise, 2000; Davis, Ford, Kherif, & Johnsrude, 2011; Rodd, Davis, & Johnsrude, 2005). It is also possible to contrast responses to a single set of stimuli while manipulating the behavioural response required (e.g., making a semantic or phonological judgement to the same set of words, Poldrack et al., 1999). Alternatively, the experimenter can make contrasts based on the listeners' performance, for example by comparing trials on which the speech was accurately perceived to trials in which it was not (e.g., Vaden Jr et al., 2013). In all these cases, the outcome measure (i.e. the dependent variable) is typically a measure of the magnitude, timing, spatial-location or spatio-temporal pattern of neural activity. In some cases, the independent variable reflects longer-term variation in language experience (e.g. comparing monolingual vs. bilingual listeners). In these cases, the outcome variable to be measured by the experimenter can be either changes in participant's neural activity, or longer term changes in their brain structure (e.g. local tissue density; see Marie & Golestani, 2016, this volume; Figure 1).

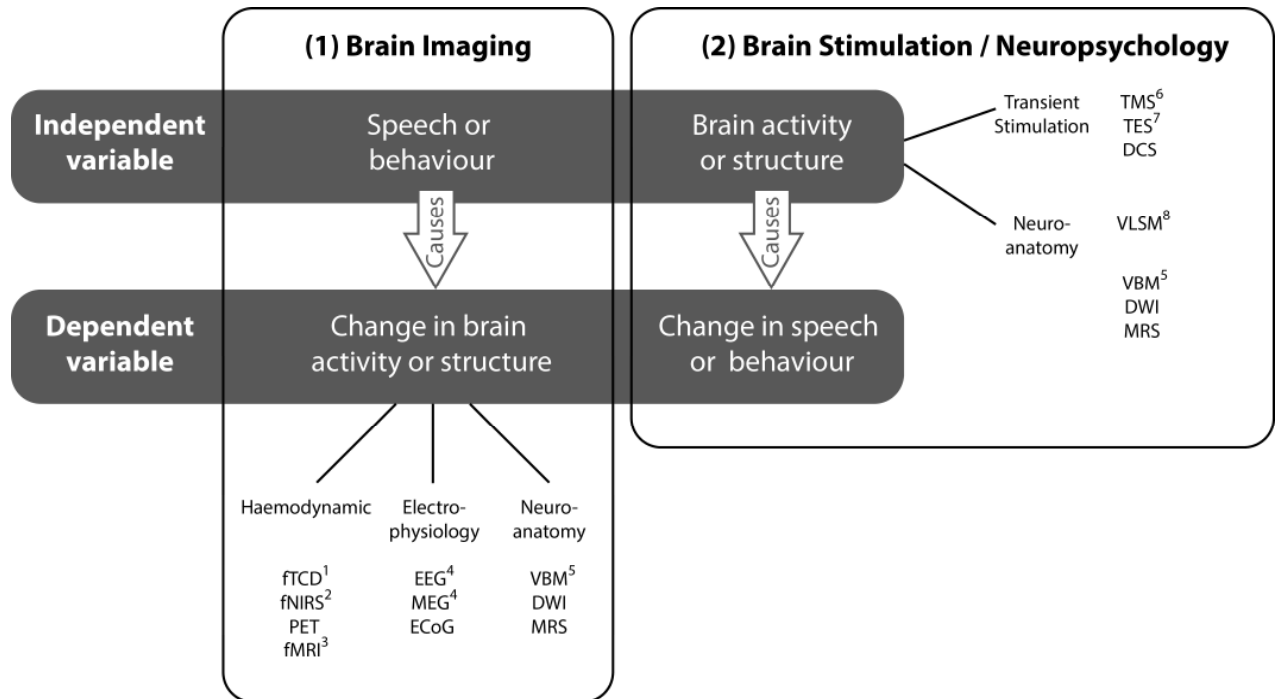


Figure 1. Taxonomy of methods for studying the neural basis of spoken language understanding. Experimental methods included in the current special issue are marked with a superscript: (1) ftCD – Functional Transcranial Doppler (Badcock & Groen, 2017, this volume), (2) fNIRS – functional Near Infrared Spectroscopy (Peelle, 2017, this volume), (3) fMRI – functional Magnetic Resonance Imaging (Evans & McGettigan, 2017, this volume), (4) EEG and MEG – Electroencephalography and Magnetoencephalography, (Wöstmann. et al., 2016, this volume), (5) Voxel-Based Morphometry (Marie & Golestani, 2016, this volume), (6) TMS – Transcranial Magnetic Stimulation (Adank et al., 2016, this volume), (7) TES – Transcranial Electrical Stimulation (Zoefel & Davis, 2016, this volume), (8) VLSM – Voxel Lesion Symptom Mapping (Wilson, 2016, this volume). Several neuroanatomical methods are listed twice in this figure to reflect uncertainty about whether neural differences are caused by or a cause of differences in behaviour. Other methods listed in the figure include: PET – Positron Emission Tomography, ECoG – Electro-corticography, DWI – Diffusion Weighted Imaging, MRS – Magnetic Resonance Spectroscopy, DCS – Direct Cortical Stimulation.

In all these cases, brain imaging experiments are “correlational” – they show changes in neural activity or structure that are a *consequence* of changes in listening conditions or listening outcomes. From these associations, it can be hard to be certain that the neural differences observed are *necessary* to support specific cognitive functions involved in speech comprehension. Many different behaviours (including non-language tasks) may activate a common set of neural regions and so any “reverse inference” that activity in a specific region supports some specific language function may be problematic (Poldrack, 2006). Despite this caveat, though it is still safe to conclude that different experimental conditions “cause” differences in brain activity (Weber & Thompson-Schill, 2010). Thus, functional imaging results can provide a sound basis for theorising about the neural basis of speech understanding and these are currently the most common methods used to explore the neural basis of spoken language understanding. This special issue will review the contributions of several different brain imaging methods. We will briefly distinguish these here by

considering three different types of neural measures: haemodynamic, electrophysiological and structural measures (see Figure 1) and refer to papers in the special issue for additional details. Having briefly surveyed these methods we will then illustrate the complementary approach adopted by experimental methods used in neuropsychological and brain stimulation.

Many of the best known methods for imaging brain activity use haemodynamic dependent measures – that is, measuring changes in blood flow and/or oxygenation that are induced by changes in neural activity rather than measuring neural activity directly. In some of the earliest forms of haemodynamic brain imaging — such as in Positron Emission Tomography (PET), and functional transcranial Doppler (ftCD; Badcock & Groen, 2017, this volume) — the dependent measure directly quantifies the rate of blood flow observed in a region or blood vessel. Blood flow measures have the advantage of being absolute physiological measures that can be directly compared between different hemispheres, individuals or experiments. However,

researchers also use other haemodynamic measures that offer superior spatial or temporal resolution, at the expense of measuring signals (such as the ratio of oxygenated and deoxygenated blood) that are a less direct measure of blood flow.

Probably the best known of these haemodynamic methods is functional Magnetic Resonance Imaging (fMRI: Evans & McGettigan, 2017, this volume) in which whole brain images of blood oxygenation can be acquired with high spatial resolution (voxel dimensions of 3 mm or less are common), but with a relatively low temporal sampling rate (typically one image every 2 seconds). However, an alternative method – functional Near Infrared Spectroscopy (fNIRS: Peelle, this volume), provides a different trade-off with a superior temporal resolution (tens of measurements per second), but correspondingly lower-spatial resolution (~10 mm, depending on the number of emitters/sensors used). While the advantages of NIRS have yet to be fully realised, these two methods in many ways provide comparable information – with fNIRS sometimes being favoured for populations (such as very young children) who may find an MRI scanner aversive, or for tasks (such as speech comprehension) in which minimising background noise during acquisition may be critical.

A different set of neural measures are obtained using electrophysiological methods such as electro- and magneto-encephalography (EEG or MEG; see Wöstmann, et al., 2016, this volume). Rather than measuring the haemodynamic consequences of neural activity these methods measure neural activity directly by recording electrical or magnetic field potentials generated by activity in large numbers of neurons. EEG and MEG measures are obtained using electrodes placed directly onto the scalp (EEG) or superconducting sensors mounted inside a close-fitting helmet (MEG). Both these methods provide excellent temporal resolution for measuring neural activity (at a millisecond time scale) at the expense of providing relatively coarse spatial information (a spatial resolution of up to ~10 mm). While the signals measured by EEG and MEG are obtained from different sensors, they provide largely common information about underlying electrical activity in the brain. More detailed spatial information about the time-course of neural activity is hard to obtain by other means in humans except by invasive implanting of grids or strips of electrodes inside the skull during neurosurgery, (ECoG; Hill et al., 2012). As explored in detail in the chapter by Wöstmann et al., (2016, this volume) key aspects of both EEG and MEG methods concern whether and how neural responses are time-aligned to cognitive or acoustic events in speech, and whether neural activity is phase-locked to these events or not (determining whether averaging of raw signals or time-frequency representations over trials is more appropriate). This methodological issue connects very directly to questions concerning

whether and how to align cognitive and neural events during speech comprehension as discussed in the previous section.

For all the neuroimaging methods considered so far, the experimenter manipulates either the stimuli or task and then observes change in neural activity that are caused by this manipulation. From a casual perspective, however, changes to some of the dependent measures provided by brain imaging may not always be a consequence of these experimental manipulations. One salient example, comes from studies in which neuroanatomical measures (i.e. differences in brain structure) are used as a dependent measure. For example, voxel-based morphometry (VBM) can be used to assess the relationship between performance on speech perception/comprehension tasks and structural properties of healthy brains (see Marie & Golestani, 2016, this volume). The specific aspect of behaviour that is tested may determine whether observed structural differences are a plausible consequence of the experimental manipulation or are more likely to be a pre-existing cause of differences in behaviour.

We will illustrate this uncertainty about behavioural and neural causes and consequences with two example studies. The first of these comes from Mechelli et al. (2004) who showed differences in neural tissue density in left inferior parietal cortex between monolingual and bilingual participants. On the assumption that the only difference between these participants was exposure to and use of a second language, this study leads us to conclude that differences in language experience cause changes in brain structure. This interpretation that behaviour (language exposure) causes neural changes is supported by a further finding from Mechelli and colleagues that structural changes in this inferior parietal region are correlated with the age at which individuals first learned their second language (greater changes following earlier acquisition). Thus, it seems likely that – in the absence of other differences between the monolingual and bilingual groups – neuro-anatomical differences are caused by differences in language experience (i.e. differences in behaviour).

A second study by Golestani, Paus & Zatorre (2002) examined the relationship between brain anatomy and the ability of English speaking participants to learn a non-native speech sound contrast (the dental/retroflex contrast used in Hindi and Urdu). They showed that the density of grey and white matter in a medial region of the left parietal lobe was correlated with individuals' abilities at acquiring this novel speech contrast. For this experiment it is implausible that success at this novel speech perception task caused a measurable change in brain structure (we would expect the same result irrespective of whether behaviour was tested before

or after MRI data was acquired). Rather, we should draw the reverse inference that differences in speech processing ability arise as a *consequence* of naturally occurring neuroanatomical variation within the population. While the ultimate cause of naturally occurring neural variation remains unclear we therefore infer that studies like that reported by Golestani et al (2002) are more appropriately grouped with those using neuropsychological patients or brain stimulation to explore how neural structure or activity causes changes in behaviour.

Studies that explore the behavioural relevance of neuroanatomical variation within the healthy population can use a range of different anatomical measures including the volume, density, thickness or shape of specific cortical and sub-cortical grey matter structures (assessed from structural MR images, as in VBM studies) or measured parameters (shape, thickness, water diffusivity) of the white matter tracts that link cortical areas (as assessed using diffusion tensor imaging and related approaches). In addition to these structural measures, other neural measures are increasingly being correlated with behavioural outcome measures in a similar way. For example, a few studies have begun to relate neurotransmitter concentrations measured using Magnetic Resonance Spectroscopy (MRS) to behavioural outcomes, e.g. in linking GABA concentration to abilities in decision making or reading (Sumner, Edden, Bompas, Evans, & Singh, 2010; Pugh et al., 2014). These methods are not yet ‘voxel based’, as spatial resolution and acquisition time is such that data is typically acquired from a single, large voxel (covering several cubic cm of cortex). However, these are further illustrations of the way in which relatively stable measures of brain structure and function can contribute to our understanding the neural foundations of spoken language understanding.

In contrast to these neuroanatomical studies of healthy controls, in which the causal relationship(s) between changes in behaviour and changes can sometimes be difficult to disentangle, studies of neuropsychological patients with speech perception/comprehension difficulties are more straightforward from a causal perspective. Neuropsychological studies routinely treat brain structure and/or function as the independent variable and use behavioural measures as dependent variables to determine the functional consequence of specific changes to neural function. Together with brain stimulation studies, neuropsychological methods are often referred to as “causal” methods since they permit a relatively strong inference that the brain region or regions that are perturbed are causally-linked to changes in behavioural outcomes. The clearest example of this neuropsychological method comes from lesion-based neuropsychology. Broca’s classic observation that a patient with damage to the left inferior frontal gyrus was unable to produce

speech (see Amunts & Zilles, 2012 for historical overview) supports the inference that this brain region is (in some way) necessary for speech production. One limitation of the traditional lesion method is that it only permits a limited degree of spatial specificity – patients with damage to Broca’s area might also have damage to many other, adjacent brain regions as well as underlying white-matter tracts. Despite dramatic improvements in structural imaging methods it can still be difficult to specify which of several co-occurring forms of damage is most responsible for differences in observed behaviour (Price, Hope, & Seghier, 2017). Nonetheless, by using MRI or CT imaging to characterise brain lesions and adopting voxel-based statistical methods, it is possible to link the specific location and extent of neural damage to functional outcomes (i.e. patterns of comprehension impairment, see Bates et al., 2003). The application of these lesion symptom mapping methods (e.g., voxel-based lesion-symptom mapping, VLSM; voxel-based morphometry, VBM), to spoken language understanding is reviewed in a paper by Wilson (this volume).

A similar form of causal inference can derive from experimentally-induced changes to brain function. Techniques for short-term stimulation of specific brain tissue allow neuropsychological methods to be used in exploring the neural basis of spoken language understanding in healthy individuals. Typical experimental designs involve choosing one or more brain regions to stimulate (as an independent variable), and exploring the impact of this stimulation on behavioural measures of speech understanding (dependent variables). Two forms of transcranial brain stimulation are reviewed in this special issue. The first of these, transcranial magnetic stimulation (TMS: Adank, Nuttall, & Kennedy-Higgins, 2016, this volume) involves magnetically inducing transient neural activity (action potentials or spikes) in cortical regions beneath an electro-magnetic coil. TMS-induced neural spiking disrupts ongoing neural activity on a short-term basis (lasting milliseconds), or (if applied repeatedly) can suppress neural activity for a longer period (tens of minutes). A second, complimentary technique, transcranial electrical stimulation, uses electrical currents applied directly to the scalp (TES: Zoefel & Davis, 2016, this volume). In contrast to TMS, TES (at comfortable levels) does not directly induce spiking activity, but can change the polarisation of neural tissues to enhance or suppress stimulus or behaviourally-evoked activity. Brain stimulation with either TMS or TES can support causal inferences similar to those allowed by lesion-based neuropsychological methods; i.e. that the stimulated brain regions contribute to a specific cognitive function or behaviour. However, these brain stimulation methods differ with respect to their regional specificity – TMS leads to more focal neural effects that can be localised to specific brain areas,

whereas TES often produces more diffuse effects (though see Datta et al., (2009) for a technique for improving the spatial precision of stimulation). They also differ with respect to functional outcomes – TMS is used primarily to disrupt neural processing, whereas TES may (in some cases) enhance neural processing. Thus, these methods can provide complementary, causal evidence concerning the neural basis of spoken language understanding in healthy individuals.

#### 4. Future directions

In looking back at the 1996 special issue, it is clear how rapidly the neuroscience of spoken language understanding has developed in the past 20 years. Few if any of the techniques explored in the present special issue were well established in 1996, and even those that were available had only limited applications to speech. For example, visual and motor fMRI responses were first reported in 1992 (Bandettini, 2012) yet there were few fMRI findings concerning the neural basis of speech understanding published before 1996 (see Price, 2012 for a review). The same is true for many of the other methods reported. Looking forward a further 20 years it is not clear whether we should expect similarly dramatic advances in the methods available to the neuroscience of spoken language understanding. Increases to the spatial resolution of brain imaging measures would be welcome, particularly for studies in which it is the fine-grained pattern of neural activity (rather than the overall magnitude or spatial location) that is used as a dependent measure (i.e. multivariate pattern analysis methods, see Evans & McGettigan, 2017 for discussion). We therefore look with interest towards developments in ultra-high field MRI (e.g. using 7 T magnets) that can enhance the spatial resolution of fMRI to the sub-mm spatial scale required for differentiating cortical laminae (e.g. Muckli, 2015; Kok et al, 2016). New types of MEG sensor - e.g. using higher temperature super-conducting sensor arrays (e.g., Chesca, John, & Mellor, 2015) that can be placed closer to the scalp – would similarly be helpful in improving the spatial resolution of electrophysiological methods. Looking at brain stimulation, ways to increase the neural specificity or to extend the reach of non-invasive brain stimulation methods (e.g. subcortical stimulation) or to better coordinate stimulation of anatomically distant, but functionally connected regions would also be of great benefit.

However, even without crystal ball gazing there are several ways in which we expect existing methods to develop that are already apparent in the published literature. The first is that multiple methods can be combined in a single study. This is most clearly seen in brain imaging studies that, as described in Figure 1, have thus far mostly focussed on collecting only one of three kinds of dependent measure (haemodynamic, electrophysiological or structural).

Each of these measures alone contributes different evidence concerning the organisation and function of neural systems supporting spoken language understanding. However, by combining multiple measures in a single study we can better understand the relationship between individual dependent measures. For example, Pelle, Troiani, Grossman & Wingfield (2011) combined VBM and fMRI to show that age-related peripheral hearing impairment had both structural and functional impacts on cortical auditory processing. Liebenthal et al (2010) showed how training in categorising non-speech sounds led to changes in both BOLD and EEG measures of neural activity in the left posterior STS. While these findings illustrate the feasibility of combining methods, relatively few studies use these combined observations to answer questions that could not have been answered in separate studies of different participants. Simultaneous collection of multimodal imaging data permits analyses in which neural measures from single trial recordings of one method (e.g. EEG) can be used to constrain or predict neural outcomes from another method (e.g. fMRI). Using variance in one type of response to guide analysis of another response provides a unique opportunity to bootstrap the spatial resolution of fMRI and temporal resolution of EEG. For example, Scheering et al (2016) use combined EEG/fMRI to show the laminar specific origin of oscillatory EEG responses (e.g. that gamma-band EEG is linked to BOLD responses in superficial cortical lamina); thereby replicating in human cortex observations that could previously only have been obtained from invasive methods. Yet, these single trial analyses are challenging given the low signal-to-noise ratio of simultaneously acquired multimodal imaging data.

Another way to combine methods is to use neuropsychological and brain imaging methods in parallel. This approach has been most apparent in functional imaging studies of brain injured populations – exploring neural activity associated with successful language function after left-hemisphere language regions have been lesioned (Price & Friston, 1999; Saur et al., 2006; Crinion & Price, 2005). Combinations of brain imaging and brain stimulation have also been demonstrated (e.g., fMRI and TMS, Ruff et al., 2006; TMS and EEG, Romei et al., 2008; Thut & Miniussi, 2009). These combined methods offer the potential to show how stimulation of specific neural systems produces behavioural impairment as neural effects of stimulation propagate through functional networks (Hallam, Whitney, Hymers, Gouws, & Jefferies, 2016). However, combining brain imaging and brain stimuli is not only technically challenging – stimulation methods often generate image artefacts that can be difficult to remove – but also leads to difficulties of interpretation. It may be unclear – particularly when using slow haemodynamic methods – which neural effects are directly related to neural stimulation,



which are linked to impaired behavioural outcomes and which are downstream consequences or compensation for more effortful or error prone performance. These challenges can be compounded for complex stimuli such as speech that engage widely distributed brain responses.

These difficulties of interpretation reflect, to our mind, another challenge that is apparent in the neuroscientific literature on spoken language understanding. At the time of the last special issue, there was a widespread acceptance that implemented computational theories – particularly in the form of connectionist models or neural network simulations – were essential to ensure that behavioural data can correctly direct theory development. The path from theory to behaviour is seldom sufficiently straightforward for verbal theories to be adequately falsified by behavioural experiments. Indeed, in the mid to late 1990s, computational models of spoken and written word recognition flourished in parallel with the experimental methods for testing these models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Norris, 1994; Gaskell & Marslen-Wilson, 1997). However, in the intervening 20 years, development of these computational theories has slowed; it is as if the scientific and technical challenges of collecting and interpreting neural data has taken scientists with computational skills away from modelling and into brain imaging. This is literally true for the present authors – we both worked on computational models of spoken and written word understanding during our PhDs (Rodd, Gaskell, & Marslen-Wilson, 2004; Davis, 2003) and subsequently moved into neuroscience.

At present, however, there is relatively little work linking new forms of neural data to computational models of spoken language (though see Ueno, Saito, Rogers, & Lambon Ralph, 2011 for an attempt in the domain of neuropsychology; Blank & Davis, 2016 in brain imaging; Tourville & Guenther, 2011 in speech production). Instead, theoretical accounts of speech processing that seek to explain neural data have largely been in the form of box and arrow drawings of functional pathways accompanied by verbal descriptions of underlying mechanisms (e.g., Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Henson, 2005). It was apparent to cognitive scientists many years ago that these verbal theories were inadequate explanations of underlying cognitive mechanisms. It should be similarly apparent to neuroscientists that verbal theories cannot substitute for a fully-implemented computational models in explaining neural data (see Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017 for similar arguments).

The future direction that we would therefore most strongly encourage for the cognitive neuroscience of spoken language understanding is for better integration of behavioural, cognitive and neural data in the form of implemented neuro-computational

models. While one might naturally hope that these models could build on the successes of existing computational theories, we acknowledge that existing models are in many cases insufficiently neural. Their components need to be mapped onto anatomical networks in the brain, and we need to develop linking hypotheses such that the same model can be used to predict many different forms of neural data (haemodynamic, electrophysiological, lesions, etc.). These linking hypothesis should in turn be founded on a detailed understanding of the underlying neurophysiology. Much work lies ahead in delivering on this promise and we would hope that a new special issue 20 years from now might lay the groundwork for adequately integrating behavioural, neural and computational theorising in the domain of spoken language understanding.

## Acknowledgements

We would like to thank Jonathan Peelle and Stephen Wilson for their helpful comments and suggestions on an earlier draft of this paper. This work was supported by the UK Economic and Social Research Council (JR: ES/K013351/1) and by the UK Medical Research Council (MHD: MC-A060-5PQ80). Both authors contributed equally to this paper and to editing this special issue.

## Keywords

Spoken Language; Neuroimaging; Cognitive Neuroscience

## Footnotes

- (1) However, just because we can measure neural responses to speech in the absence of secondary tasks, this does not mean that behavioural measures should be excluded from neuroimaging studies. For example, in one fMRI study of speech comprehension we observed largely identical neural responses to high versus low ambiguity sentences during in the absence and presence of an engaging comprehension task (Rodd, Davis & Johnsrude, 2005). Yet, we also observed greater variability in the neural responses observed during passive listening that are plausibly due to inattentive participants being less engaged in the comprehension process. (See Sabri et al., 2008; Wild et al., 2012 for further studies of these attentional effects). More generally, we seek mechanistic theories that explain the links between neural responses and behavioural outcomes; these theories must therefore explain participants' behaviour during active tasks (see Henson, 2005; Taylor, Rastle, & Davis, 2014 for discussion).

## References

- Adank, P., Nuttall, H. E., & Kennedy-Higgins, D. Transcranial magnetic stimulation and motor evoked potentials in speech perception research. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1257816
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 13367-13372. DOI: 10.1073/pnas.201400998
- Amunts, K. & Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends in Cognitive Sciences*, *16*, 418-426. DOI: 10.1016/j.tics.2012.06.005
- Badcock, N. A. & Groen, M. A. What can functional Transcranial Doppler Ultrasonography tell us about spoken language understanding? *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1276608
- Bandettini, P. A. (2012). Twenty years of functional MRI: The science and the stories. *NeuroImage*, *62*, 575-588. DOI: 10.1016/j.neuroimage.2012.04.026
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T. et al. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, *6*, 448-450. DOI: 10.1038/nn1050
- Blank, H. & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLoS Biology*, *14*. DOI: 10.1371/journal.pbio.1002577
- Chesca, B., John, D., & Mellor, C. J. (2015). Flux-coherent series SQUID array magnetometers operating above 77 K with superior white flux noise than single-SQUIDs at 4.2 K. *Applied Physics Letters*, *107*. DOI: 10.1063/1.4932969
- Coleman, M. R., Davis, M. H., Rodd, J. M., Robson, T., Ali, A., Owen, A. M. et al. (2009). Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain*, *132*, 2541-2552. DOI: 10.1093/brain/awp183
- Coleman, M. R., Rodd, J. M., Davis, M. H., Johnsrude, I. S., Menon, D. K., Pickard, J. D. et al. (2007). Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain*, *130*, 2494-2507. DOI: 10.1093/brain/awm170
- Crinion, J. & Price, C. J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*, *128*, 2858-2871. DOI: 10.1093/brain/awh659
- Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., & Bikson, M. (2009). Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimulation*, *2*, 201-207. DOI: 10.1016/j.brs.2009.03.005
- Davis, M. H. (2003). Connectionist modelling of lexical segmentation and vocabulary acquisition. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (Hove, UK: Psychology Press).
- Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F. et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proc.Natl.Acad.Sci.U.S.A*, *104*, 16032-16037. DOI: 10.1073/pnas.0701309104
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*, 3914-3932. DOI: 10.1162/jocn\_a\_00084
- Davis, M. H. & Rodd, J. M. (2011). Brain structures underlying lexical processing of speech: Evidence from brain imaging. In G. Gaskell & P. Zwitserlood (Eds.), *Lexical representation: A multidisciplinary approach* (pp. 197-230). Berlin: Mouton de Gruyter.
- Devlin, J. T. & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain*, *130*, 610-622. DOI: 10.1093/brain/awl331
- Dhamne, S. C., Kothare, R. S., Yu, C., Hsieh, T. H., Anastasio, E. M., Oberman, L. et al. (2014). A measure of acoustic noise generated from transcranial magnetic stimulation coils. *Brain Stimulation*, *7*, 432-434. DOI: 10.1016/j.brs.2014.01.056
- Evans, S. & McGettigan, C. Comprehending auditory speech: previous and potential contributions of functional MRI. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1272703
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*, 615-621. DOI: 10.1016/j.cub.2012.02.015
- Gaskell, M. G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613-656. DOI: 10.1080/016909697386646
- Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, *11*, 559-567. DOI: 10.1080/016909696386944
- Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, *35*, 997-1010. DOI:10.1016/S0896-6273(02)00862-0
- Grosjean, F. (1996). Gating. *Language and Cognitive Processes*, *11*, 597-604. DOI: 10.1080/016909696386999
- Grosjean, F. & Frauenfelder, U. H. (1996). A Guide to Spoken Word Recognition Paradigms: Introduction. *Language and Cognitive Processes*, *11*, 553-558. DOI: 10.1080/016909696386935
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R. et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain*

- Mapping*, 7, 213-223. DOI: 10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N
- Hallam, G. P., Whitney, C., Hymers, M., Gouws, A. D., & Jefferies, E. (2016). Charting the effects of TMS with fMRI: Modulation of cortical recruitment within the distributed network supporting semantic control. *Neuropsychologia*, 93, 40-52. DOI: 10.1016/j.neuropsychologia.2016.09.012
- Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48, 865-892. DOI: 10.1515/LING.2010.027
- Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 58, 193-233. DOI: 10.1080/02724980443000502
- Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393-402. DOI: 10.1038/nrn2113
- Hill, N. J., Gupta, D., Brunner, P., Gunduz, A., Adamo, M. A., Ritaccio, A. et al. (2012). Recording human electrocorticographic (ECoG) signals for neuroscientific research and real-time functional cortical mapping. *Journal of visualized experiments*. DOI: 10.3791/3993
- Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology*, 19, 51-65. DOI: 10.1044/1058-0360(2009/08-0083)
- Kilborn, K. & Moss, H. (1996). Word monitoring. *Language and Cognitive Processes*, 11, 689-694. DOI: 10.1080/016909696387105
- Kocagoncu, E., Clarke, A., Devereux, B. J., & Tyler, L. K. (2017). Decoding the cortical dynamics of sound-meaning mapping. *Journal of Neuroscience*, 37, 1312-1319. DOI: 10.1523/JNEUROSCI.2858-16.2016
- Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2016). Neurobiological Basis of Language Learning Difficulties. *Trends in Cognitive Sciences*, 20, 701-714. DOI: 10.1016/j.tics.2016.06.012
- Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, 111, 2433-2444. DOI: 10.1152/jn.00497.2013
- Liebethal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, 20, 2958-2970. DOI: 10.1093/cercor/bhq045
- MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, 3. DOI: 10.1038/ncomms1715
- Marie, D. & Golestani, N. Brain structural imaging of receptive speech and beyond: a review of current methods. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1250926
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189, 226-228. DOI: 10.1126/science.189.4198.226
- Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 125-150). Hillsdale NJ: Erlbaum.
- Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522-523. DOI:10.1038/244522a0
- Martin, C. D., Garcia, X., Potter, D., Melinger, A., & Costa, A. (2016). Holiday or vacation? The processing of variation in vocabulary across dialects. *Language, Cognition and Neuroscience*, 31, 375-390. DOI: 10.1080/23273798.2015.1100750
- Mechelli, A., Crinion, J. T., Noppeney, U., O' Doherty, J., Ashburner, J., Frackowiak, R. S. et al. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature*, 431, 757. DOI: 10.1038/431757a
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378. DOI: 10.1121/1.397688
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189-234. DOI: 10.1016/0010-0277(94)90043-4
- Nygaard, L. C. & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, 60, 355-376. DOI: 10.3758/BF03206860
- O'Rourke, T. B. & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, 60, 121-150. DOI: 10.1016/S0301-0511(02)00045-5
- Passingham, R. E. & Rowe, J. B. (2015). *A Short Guide to Brain Imaging: The Neuroscience of Human Cognition*. Oxford University Press.
- Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, 8. 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, 1410-1419. DOI: 10.1016/j.neuroimage.2010.05.015
- Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *Journal of Neuroscience*, 31, 12638-12643. DOI: 10.1523/JNEUROSCI.2559-11.2011
- Perrachione, T. H. & Ghosh, S. S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. *Frontiers in Neuroscience*, 7. DOI: 10.3389/fnins.2013.00055

- Pisoni, D. B. (1996). Word identification in noise. *Language and Cognitive Processes, 11*, 681-687. DOI: 10.1080/016909696387097
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115. DOI: 10.1037//0033-295X.103.1.56
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*, 59-63. DOI: 10.1016/j.tics.2005.12.004
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage, 10*, 15-35. DOI: 10.1006/nimg.1999.0441
- 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*, 816-847. DOI: 10.1016/j.neuroimage.2012.04.062
- Price, C. J. & Friston, K. J. (1999). Scanning patients with tasks they can perform. *Human Brain Mapping, 8*, 102-108. DOI: 10.1002/(SICI)1097-0193(1999)8:2/3<102::AID-HBM6>3.0.CO;2-J
- Price, C. J., Hope, T. M., & Seghier, M. L. (2017). Ten problems and solutions when predicting individual outcome from lesion site after stroke. *NeuroImage, 145*, 200-208. DOI: 10.1016/j.neuroimage.2016.08.006
- Pugh, K. R., Frost, S. J., Rothman, D. L., Hoeft, F., Del Tufo, S. N., Mason, G. F. et al. (2014). Glutamate and choline levels predict individual differences in reading ability in emergent readers. *Journal of Neuroscience, 34*, 4082-4089. DOI: 10.1523/JNEUROSCI.3907-13.2014
- Rauschecker, J. P. & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*, 718-724. DOI: 10.1038/nn.2331
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*, 1261-1269. DOI: 10.1093/cercor/bhi009
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*, 89-104. DOI: 10.1016/j.cogsci.2003.08.002
- Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language, 141*, 89-102. DOI: 10.1016/j.bandl.2014.11.012
- Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., & Thut, G. (2008). Spontaneous fluctuations in posterior  $\alpha$ -band EEG activity reflect variability in excitability of human visual areas. *Cerebral Cortex, 18*, 2010-2018. DOI: 10.1093/cercor/bhm229
- Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J. D. et al. (2006). Concurrent TMS-fMRI and Psychophysics Reveal Frontal Influences on Human Retinotopic Visual Cortex. *Current Biology, 16*, 1479-1488. DOI: 10.1016/j.cub.2006.06.057
- Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitel, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *NeuroImage, 39*, 1444-1456. DOI: 10.1016/j.neuroimage.2007.09.052
- Saur, D. & Hartwigsen, G. (2012). Neurobiology of language recovery after stroke: Lessons from neuroimaging studies. *Archives of Physical Medicine and Rehabilitation, 93*, S15-S25. DOI: 10.1016/j.apmr.2011.03.036
- Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M. et al. (2006). Dynamics of language reorganization after stroke. *Brain, 129*, 1371-1384. DOI: 10.1093/brain/awl090
- Scheering, R., Koopmans, P. J., Van Mourik, T., Jensen, O., & Norris, D. G. (2016). The relationship between oscillatory EEG activity and the laminar-specific BOLD signal. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 6761-6766. DOI: 10.1073/pnas.1522577113
- 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*, 774-782. DOI: 10.1016/j.neuroimage.2005.08.025
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*, 2400-6. DOI: 10.1093/brain/123.12.2400
- Sumner, P., Edden, R. A. E., Bompas, A., Evans, C. J., & Singh, K. D. (2010). More GABA, less distraction: A neurochemical predictor of motor decision speed. *Nature Neuroscience, 13*, 825-827. DOI: 10.1038/nn.2559
- Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes, 11*, 569-576. DOI: 10.1080/016909696386953
- Tanenhaus, M. K. & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes, 11*, 583-588. DOI: 10.1080/016909696386971
- Taylor, J. S. H., Rastle, K., & Davis, M. H. (2014). Interpreting response time effects in functional imaging studies. *NeuroImage, 99*, 419-433. DOI: 10.1016/j.neuroimage.2014.05.073
- Thut, G. & Miniussi, C. (2009). New insights into rhythmic brain activity from TMS-EEG studies. *Trends in Cognitive Sciences, 13*, 182-189. DOI: 10.1016/j.tics.2009.01.004
- Tourville, J. A. & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*, 952-981. DOI: 10.1080/01690960903498424
- Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of*

- Mathematical Psychology*, 76, 65-79. DOI: 10.1016/j.jmp.2016.01.001
- Ueno, T., Saito, S., Rogers, T., & Lambon Ralph, M. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72, 385-396. DOI: 10.1016/j.neuron.2011.09.013
- Vaden Jr, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal of Neuroscience*, 33, 18979-18986. DOI: 10.1523/JNEUROSCI.1417-13.2013
- Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., & Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. *Journal of Neuroscience*, 32, 9089-9102. DOI: 10.1523/JNEUROSCI.5685-11.2012
- Van Berkum, J. J. A., Van Den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580-591. DOI: 10.1162/jocn.2008.20054
- Weber, M. J. & Thompson-Schill, S. L. (2010). Functional neuroimaging can support causal claims about brain function. *Journal of Cognitive Neuroscience*, 22, 2415-2416. DOI: 10.1162/jocn.2010.21461
- 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, 14010-14021. DOI: 10.1523/JNEUROSCI.1528-12.2012
- Wilson, S. M. Lesion-symptom mapping in the study of spoken language understanding. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1248984
- Wöstmann., Fiedler, L., & Obleser, J. Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1262051
- Zoefel, B. & Davis, M. H. Transcranial electric stimulation for the investigation of speech perception and comprehension. *Language, Cognition and Neuroscience*, (in press). DOI: 10.1080/23273798.2016.1247970
- Zwitserslood, P. (1996). Form priming. *Language and Cognitive Processes*, 11, 589-596. DOI: 10.1080/016909696386980
- Zwitserslood, P. & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, 10, 121-136. DOI: 10.1080/01690969508407090