OPINION

The cortical organization of speech processing

Gregory Hickok and David Poeppel

Abstract | Despite decades of research, the functional neuroanatomy of speech processing has been difficult to characterize. A major impediment to progress may have been the failure to consider task effects when mapping speech-related processing systems. We outline a dual-stream model of speech processing that remedies this situation. In this model, a ventral stream processes speech signals for comprehension, and a dorsal stream maps acoustic speech signals to frontal lobe articulatory networks. The model assumes that the ventral stream is largely bilaterally organized — although there are important computational differences between the left- and right-hemisphere systems — and that the dorsal stream is strongly left-hemisphere dominant.

Understanding the functional anatomy of speech perception has been a topic of intense investigation for more than 130 years, and interest in the basis of speech and language dates back to the earliest recorded medical writings. Despite this attention, the neural organization of speech perception has been surprisingly difficult to characterize, even in gross anatomical terms.

The first hypothesis, dating back to the 1870s[1], was straightforward and intuitive: speech perception is supported by the auditory cortex. Evidence for this claim came from patients with auditory language comprehension disorders (today's Wernicke's aphasics) who typically had lesions of the left superior temporal gyrus (STG). Accordingly, the left STG in particular was thought to support speech perception. This position was challenged by two discoveries in the 1970s and 1980s. The first was that deficits in the ability to perceive speech sounds contributed minimally to the auditory comprehension deficit in Wernicke's aphasia[2–7]. The second was that destruction of the left STG does not lead to deficits in the auditory comprehension of speech, but instead causes deficits in speech production[8]. These findings do not rule out a role for the left STG in speech perception, but make it clear that additional regions participate in the process.

Around the same time, neuropsychological experiments showed that damage to frontal or inferior parietal areas in the left hemisphere caused deficits in tasks that required the discrimination or identification of speech syllables[3,4,9]. These findings highlighted the possible role of a fronto-parietal circuit in the perception of speech. However, the relevance of these findings in mapping the neural circuits for speech perception was questionable, given that the ability to perform syllable discrimination and identification tasks doubly dissociated from the ability to comprehend aurally presented words: that is, there are patients who have impaired syllable discrimination but good word comprehension, and vice versa[7,10]. Some authors took this to indicate that auditory comprehension deficits in Wernicke's aphasia that are secondary to left temporal lobe lesions resulted from disruptions of semantic rather than phonological processes[11], whereas others postulated that the mapping between phonological and semantic representations was disrupted[10]. Proposed explanations for deficits in syllable discrimination associated with non-temporal lobe lesions included generalized attentional deficits[11] and phonological working memory deficits[3].

The clinical picture regarding the neural basis of speech perception was, therefore, far from clear when functional imaging methods arrived on the scene in the 1980s. Unfortunately, the first imaging studies failed to clarify the situation: studies presenting speech stimuli in passive listening tasks highlighted superior temporal regions bilaterally[12,13], and studies that used tasks similar to syllable discrimination or identification found prominent activations in the left STG[14] and left inferior frontal lobe[15]. The paradox remained: damage to either of these left-hemisphere regions primarily produced speech production deficits (or at most some mild auditory comprehension deficits affecting predominantly post-phonemic processing), not the impaired auditory comprehension problems one might expect if the main substrate for speech perception had been destroyed. Although several authors discussed the possibility that left inferior frontal areas might be activated in these studies as a result of task-related phonological working memory processes[14,16], thus addressing the paradox for one region, there was little discussion in the imaging literature of the paradox in connection with the left STG.

The goal of this article is to describe and extend a dual-stream model of speech processing that resolves this paradox. The concept of dual processing streams dates back at least to the 1870s, when Wernicke proposed his now famous model of speech processing, which distinguished between two pathways leading from the auditory system[1]. A dual-stream model has been well accepted in the visual domain since the 1980s[17], and the concept of a similar arrangement in the auditory system has gained recent empirical support[18,19]. The model described in this article builds on this earlier work. We will outline the central components and assumptions of the model and discuss the relevant evidence.

Task dependence and definitions

From the studies described above, it is clear that the neural organization of speech processing is task dependent. Therefore, when attempting to map the neural systems

Box 1 | Linguistic organization of speech

Linguistic research demonstrates that there are multiple levels of representation in mapping sound to meaning. The distinct levels conjectured to form the basis for speech include 'distinctive features', the smallest building blocks of speech that also have an acoustic interpretation; as such they provide a connection between action and perception in speech. Distinctive features provide the basic inventory characterizing the sounds of all languages[97–100]. Bundles of coordinated distinctive features overlapping in time constitute segments (often called phonemes, although the terms have different technical meanings). Languages have their own phoneme inventories, and the sequences of these are the building blocks of words[101]. Segments are organized into syllables, which have language-specific structure. Some languages permit only a few syllable types, such as consonant (C)–vowel (V), whereas others allow for complex syllable structure, such as CCVCC[102]. In recent research, syllables are proposed to be central to parsing the speech stream into manageable chunks for analysis[103]. Featural, segmental and syllabic levels of representation provide the infrastructure for prelexical phonological analysis[104,105]. The smallest building blocks mediating the representation of meaning are morphemes. Psycholinguistic and neurolinguistic evidence suggests that morphemic structure has an active role in word recognition and is not just a theoretical construct[106]. There are further levels that come into play (such as syntactic information and compositional semantics, including sentence- and discourse-level information), but those mentioned above are the representations that must be accounted for in speech perception in the service of lexical access. At that point one can make contact with existing detailed models of lexical access and representation, including versions of the cohort model[107], the neighbourhood activation model[108] and others. These provide further lexical-level (as well as prelexical) constraints to be accounted for.

supporting speech processing, one must carefully define the computational process (task) of interest. This is not always clearly done in the literature. Many studies using the term 'speech perception' to describe the process of interest employ sublexical speech tasks, such as syllable discrimination, to probe that process. In fact, speech perception is sometimes interpreted as referring to the perception of speech at the sublexical level. However, the ultimate goal of these studies is presumably to understand the neural processes supporting the ability to process speech sounds under ecologically valid conditions, that is, situations in which successful speech sound processing ultimately leads to contact with the mental lexicon (BOX 1) and auditory comprehension. Thus, the implicit goal of speech perception studies is to understand sublexical stages in the process of speech recognition (auditory comprehension). This is a perfectly reasonable goal, and the use of sublexical tasks would seem to be a logical choice for assessing these sublexical processes, except for the empirical observation that speech perception and speech recognition doubly dissociate. The result is that many studies of speech perception have only a tenuous connection to their implicit target of investigation, speech recognition.

In this article we use the term 'speech processing' to refer to any task involving aurally presented speech. We will use speech perception to refer to sublexical tasks (such as syllable discrimination), and speech recognition to refer to the set of computations that transform acoustic signals into a representation that makes contact with the mental lexicon. This definition of speech recognition does not require that there be only one route to lexical access: recognition could be achieved via parallel, computationally differentiated mappings. According to these definitions, and the dissociations described above, it follows that speech perception does not necessarily correlate with, or predict, speech recognition. This presumably results from the fact that the computational end points of these two abilities are distinct. We have suggested that there is overlap between these two classes of tasks in the computational operations leading up to and including the generation of sublexical representations, but that different neural systems are involved beyond this stage[6]. Speech recognition tasks involve lexical access processes, whereas speech perception tasks do not need lexical access but instead require processes that allow the listener to maintain sublexical representations in an active state during the performance of the task, as well as the recruitment of task-specific operations. Thus, speech perception tasks involve some degree of executive control and working memory, which might explain the association with frontal lobe lesions and activations[3,14,16].

Dual-stream model of speech processing

With these comments in mind, we can summarize the central claims of the dual-stream model (FIG. 1). Similar to previous hypotheses regarding the auditory 'what' stream[18], this model proposes that a ventral stream, which involves structures in the superior and middle portions of the temporal lobe, is involved in processing speech signals for comprehension (speech recognition). A dorsal stream, which involves structures in the posterior frontal lobe and the posterior dorsal-most aspect of the temporal lobe and parietal operculum, is involved in translating acoustic speech signals into articulatory representations in the frontal lobe, which is essential for speech development and normal speech production. The suggestion that the dorsal stream has an auditory–motor integration function differs from earlier arguments for a dorsal auditory 'where' system[18], but is consistent with recent conceptualizations of the dorsal visual stream (BOX 2) and has gained support in recent years[20–22]. We propose that speech perception tasks rely to a greater extent on dorsal stream circuitry, whereas speech recognition tasks rely more on ventral stream circuitry (with shared neural tissue in the left STG), thus explaining the observed double dissociations. In addition, in contrast to the typical view that speech processing is mainly left-hemisphere dependent, the model suggests that the ventral stream is bilaterally organized (although with important computational differences between the two hemispheres); so, the ventral stream itself comprises parallel processing streams. This would explain the failure to find substantial speech recognition deficits following unilateral temporal lobe damage. The dorsal stream, however, is strongly left-dominant, which explains why production deficits are prominent sequelae of dorsal temporal and frontal lesions, and why left-hemisphere injury can substantially impair performance in speech perception tasks.

Ventral stream: sound to meaning

Mapping acoustic speech input onto conceptual and semantic representations involves multiple levels of computation and representation. These levels may include the representation of distinctive features, segments (phonemes), syllabic structure, phonological word forms, grammatical features and semantic information (BOX 1). However, it is unclear whether the neural computations that underlie speech recognition in particular involve each of these processing levels, and whether the levels involved are serially organized and immutably applied or involve parallel computational pathways and allow for some flexibility in processing. These questions cannot be answered definitively given the existing evidence. However, some progress has been made in understanding the functional organization of the pathways that map between sound and meaning.

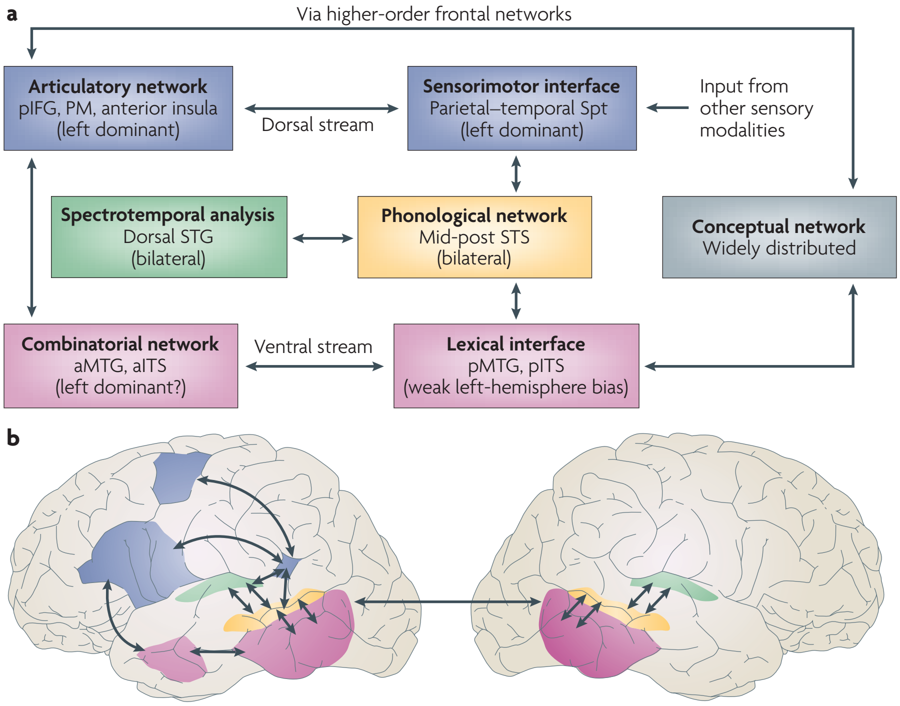

Figure 1 | **The dual-stream model of the functional anatomy of language. a** | Schematic diagram of the dual-stream model. The earliest stage of cortical speech processing involves some form of spectrotemporal analysis, which is carried out in auditory cortices bilaterally in the supratemporal plane. These spectrotemporal computations appear to differ between the two hemispheres. Phonological-level processing and representation involves the middle to posterior portions of the superior temporal sulcus (STS) bilaterally, although there may be a weak left-hemisphere bias at this level of processing. Subsequently, the system diverges into two broad streams, a dorsal pathway (blue) that maps sensory or phonological representations onto articulatory motor representations, and a ventral pathway (pink) that maps sensory or phonological representations onto lexical conceptual representations. **b** | Approximate anatomical locations of the dual-stream model components, specified as precisely as available evidence allows. Regions shaded green depict areas on the dorsal surface of the superior temporal gyrus (STG) that are proposed to be involved in spectrotemporal analysis. Regions shaded yellow in the posterior half of the STS are implicated in phonological-level processes. Regions shaded pink represent the ventral stream, which is bilaterally organized with a weak left-hemisphere bias. The more posterior regions of the ventral stream, posterior middle and inferior portions of the temporal lobes correspond to the lexical interface, which links phonological and semantic information, whereas the more anterior locations correspond to the proposed combinatorial network. Regions shaded blue represent the dorsal stream, which is strongly left dominant. The posterior region of the dorsal stream corresponds to an area in the Sylvian fissure at the parietotemporal boundary (area Spt), which is proposed to be a sensorimotor interface, whereas the more anterior locations in the frontal lobe, probably involving Broca's region and a more dorsal premotor site, correspond to portions of the articulatory network. aITS, anterior inferior temporal sulcus; aMTG, anterior middle temporal gyrus; pIFG, posterior inferior frontal gyrus; PM, premotor cortex.

Parallel computations and bilateral organization. Models of speech recognition typically assume that a single computational pathway exists. In general, the prevailing models (including the TRACE[23], cohort[24] and neighbourhood activation[25] models) assume that various stages occur in series that map from sounds to meaning, typically incorporating some acoustic-, followed by some phonetic- and finally some lexical-level representations. Although activation of representations within and across stages can be parallel and interactive, there is nonetheless only one computational route from sound to meaning. By contrast, the model we suggest here proposes that there are multiple routes to lexical access, which are implemented as parallel channels. We further propose that this system is organized bilaterally, in contrast to many neural accounts of speech processing[26–31].

From a behavioural standpoint, it is clear that the speech signal contains multiple, partially redundant spectral and temporal cues that can be exploited by listeners and that allow speech perception to tolerate a range of signal degradation conditions[32–36]. This supports the idea that redundant computational mechanisms — that is, parallel processing — might exist to exploit these cues.

There is also strong neural evidence that parallel pathways are involved in speech recognition. Specifically, evidence from patients with unilateral damage to either hemisphere, split-brain patients[37] (who have undergone sectioning of the corpus callosum) and individuals undergoing Wada procedures[38] (a presurgical procedure in which one or the other cerebral hemisphere is selectively anaesthetized to assess language and memory lateralization patterns) indicates that there is probably at least one pathway in each hemisphere that can process speech sounds sufficiently well to access the mental lexicon[5,6]. Furthermore, bilateral damage to superior temporal lobe regions is associated with severe deficits in speech recognition (word deafness)[39], consistent with the idea that speech recognition systems are bilaterally organized[40]. In rare cases, word deafness can also result from focal unilateral lesions[41]; however, the frequency of such an occurrence is exceedingly small relative to the frequency of occurrence of unilateral lesions generally, suggesting that such cases are the exception rather than the rule.

Functional imaging evidence is also consistent with bilateral organization of speech recognition processes. A consistent and uncontroversial finding is that, when contrasted with a resting baseline, listening to speech activates the STG bilaterally, including the dorsal STG and superior temporal sulcus (STS). Many studies have attempted to identify 'phonemic processing' more specifically by contrasting speech stimuli with various non-speech controls (BOX 3). Most studies find bilateral activation for speech typically in the superior temporal sulcus (STS), even after subtracting out the non-speech controls (TABLE 1). However, in many of these studies, the activation is more extensive, or, in a few studies, is solely found in the left hemisphere. Nevertheless, we do not believe that this constitutes evidence against a bilateral organization of speech perception, because a region activated by speech that is also activated by acoustically similar non-speech stimuli could still be involved in, or capable of, speech processing. Specificity is not a prerequisite for functional effectiveness. For example, the vocal tract is highly effective (even specialized) for speech production, but is far from speech-specific as it is also functionally effective for digestion. Furthermore, it is not clear exactly what is being isolated in these 'phoneme-specific' areas. It has been hard to identify differential

Box 2 | Sensorimotor integration in the parietal lobe

The dorsal stream processing system in vision was first conceptualized as subserving a spatial 'where' function[17]. However, more recent work has suggested a more general functional role in visuomotor integration[72,73,93]. Single-unit recordings in the primate parietal lobe have identified cells that are sensitive not only to visual stimulation, but also to action towards the visual stimulation. For example, a unit may respond when an object is presented, but also when the monkey reaches for that object in an appropriate way, even if the object is no longer in view[92,93]. A number of visuomotor areas have been identified that appear to be organized around different motor effector systems[93,94]. These visuomotor areas are densely connected with frontal lobe regions involved in controlling motor behaviour for the various effectors[93]. In humans, dissociations have been observed between the conscious perception of visual information (ventral stream function) and the ability to act on that information appropriately (dorsal stream function). For example, in optic ataxia, patients can judge the location and orientation of visual stimuli, but have substantial deficits in reaching for those same stimuli; parietal lesions are associated with optic ataxia[100]. The reverse dissociation has also been reported in the form of a case of visual agnosia in which perceptual judgements of object orientation and shape were severely impaired, yet reaching behaviour towards those same stimuli showed normal shape- and orientation-dependent anticipatory movements[101].

activations in response to words versus pseudowords[42], even though the latter are presumably not represented in one's mental lexicon. An accepted explanation is that pseudowords activate lexical networks via features (phonemes or syllables) that are shared with real words. Likewise, it is possible that a similar phenomenon is occurring at sublexical processing levels: non-phonemic acoustic signals might activate phonological networks because they share acoustic features with signals that contain phonemic information. Thus, phonological networks in both the left and right hemispheres might be subtracted out of many of these studies. Finally, even if left-dominant phoneme-specific networks have been isolated by these studies, it remains possible that other networks (the right hemisphere, for example) represent speech as non-categorical acoustic signals that can be used to access the mental lexicon. Thus, even taking into account the most extreme interpretation of the imaging data, the claim that the speech recognition system is bilaterally organized (accounting for the lesion data), but with important computational differences (accounting for asymmetries), remains the most viable position.

Multi-time resolution processing. Mapping from sound to an interpretable representation involves integrating information on different timescales. Determining the order of segments that constitute a lexical item (such as the difference between 'pets' and 'pest') requires information encoded in temporal windows of ~20–50 ms. To achieve successful lexical access (and to succeed on numerous other temporal order tasks in auditory perception), the input signal must be analysed at this scale. Suprasegmental information carried on the syllable occurs over longer intervals, roughly 150–300 ms. This longer scale, roughly commensurate with the acoustic envelope of a spoken utterance, carries syllable-boundary and syllabic-rate cues as well as (lexical) tonal information, prosodic cues (for interpretation) and stress cues.

Although the presupposition typically remains unstated, previous models have assumed that either hierarchical processing (a segmental analysis followed by syllable construction on the basis of smaller units)[43,44] or parsing at the syllabic rate[45,46] take place with no privileged analysis at the segmental or phonemic level.

Another way to integrate these differing requirements is to posit a multi-time resolution model in which speech is processed concurrently on these two timescales by two separate streams, and the information that is extracted is combined for subsequent computations at the lexical level[47,48] (FIG. 2). It is possible that lexical access is initiated by information from each individual stream, but that optimal lexical access occurs when segmental-rate and syllabic-rate information is combined. Determining whether such a model is correct requires evidence for short-term integration, long-term integration and perceptual interaction. Although still tentative, there is now evidence supporting each of these conjectures[47,48]. Perceptually meaningful interaction at these two timescales has been recently demonstrated, in a study showing that stimuli that selectively filter out or preserve these specific modulation frequencies lead to performance changes (D.P., unpublished observations). Functional MRI data also support the hypothesis that there is multi-time resolution processing, and that this processing is hemispherically asymmetrical, with the right hemisphere showing selectivity for long-term integration. The left hemisphere seems less selective in its response to different integration timescales[47]. This finding differs from the view that the left hemisphere is dominant for processing (fast) temporal information, and the right hemisphere is dominant for processing spectral information[49,50]. Instead, we propose that neural mechanisms for integrating information over longer timescales are predominantly located in the right hemisphere, whereas mechanisms for integrating over shorter timescales might be represented more bilaterally. Thus, we are suggesting that the traditional view of the left hemisphere

Box 3 | Is speech special?

The question of whether the machinery to analyse speech is 'special' has a contentious history in the behavioural literature[43,102–104], and has migrated into neuroimaging as an area of research[105,106]. The concept of specialization presumably means that some aspect of the substrate involved in speech analysis is dedicated, that is, optimized for the perceptual analysis of speech as opposed to other classes of acoustic information. Whether the debate has led to substantive insight remains unclear, but the issues have stimulated valuable research.

It may or may not be the case that there are populations of neurons whose response properties encode speech in a preferential manner. In central auditory areas, information in the time domain is especially salient, and perhaps some neuronal populations are optimized for dealing with temporal information commensurate with information in speech signals. These are open questions, but with little effect on the framework presented here.

There is, however, an 'endgame' for the perceptual process, and that is to make contact with lexical representations to mediate intelligibility. We propose that at the sound–word interface there must be some specialization, for the following reason: lexical representations are used for subsequent processing, entering into phonological, morphological, syntactic and compositional semantic computations. For this, the representation has to be in the correct format, which appears to be unique to speech. For example, whistles, birdsong or clapping do not enter into subsequent linguistic computation, although they receive a rich analysis. It therefore stands to reason that lexical items have some representational property that sets them apart from other auditory information. There must be a stage in speech recognition at which this format is constructed, and if that format is of a particular type, there is necessarily specialization.

being uniquely specialized for processing fast temporal information is incorrect, or at most weakly supported by existing data.

Another recently discussed possibility[26] is that the left hemisphere might be predisposed to processing or representing acoustic information more categorically than the right hemisphere. This could explain some of the asymmetries found in functional activation studies of speech perception, and might also accommodate the lesion data on the assumption that less categorical representations of speech in the right hemisphere are sufficient for lexical access.

Table 1 | **A sample of recent functional imaging studies of sublexical speech processing**

Speech stimuli	Control stimuli	Task	Coordinates			Coordinate space	Hemisphere	Ref.
			x	y	z			
CVCs	Sinewave analogues	Oddball detection	–60	–16	–8	Talairach	L	31
			–64	–36	0	Talairach	L	
			–64	–44	12	Talairach	L	
			56	–28	–4	Talairach	R	
			52	–20	–16	Talairach	R	
			52	–16	0	Talairach	R	
CVs	Tones	Target detection	–64	–12	–8	MNI	L	102
			–56	–16	–12	MNI	L	
			–60	–24	8	MNI	L	
			44	–24	8	MNI	R	
			52	–28	8	MNI	R	
CVs	Tones + noise	Passive listening	–64	–20	0	MNI	L	
			–64	–32	4	MNI	L	
			–64	–28	8	MNI	L	
			56	–12	–8	MNI	R	
			52	–20	–8	MNI	R	
CVs	Noise	Passive listening	56	–8	–4	MNI	R	
			64	–20	0	MNI	R	
			60	–24	–8	MNI	R	
			–64	–16	0	MNI	L	
			–68	–28	–4	MNI	L	
			–64	–36	4	MNI	L	
			–44	–28	12	MNI	L	
			–64	–28	8	MNI	L	
CVs	Sinewave analogues	AX discrimination	–56	–22	3	Talairach	L	103
			–51	–14	–4	Talairach	L	
			54	–46	9	Talairach	R	
			52	–25	2	Talairach	R	
			62	1	–12	Talairach	R	
Sinewave CVs perceived as speech	Sinewave CVs perceived as non-speech	Oddball detection	–56	–40	0	Talairach	L	101
			–60	–24	4	Talairach	L	
Synthesized CV continuum	Spectrally rotated synthesized CVs	ABX discrimination	–60	–8	–3	Talairach	L	26
			–56	–31	3	Talairach	L	
CVs	Noise	Detect repeating sounds	–59	–27	–2	MNI	L	104
			–63	–16	–6	MNI	L	
			59	–4	–10	MNI	R	
Sinewave CVCs	Sinewave non-speech analogues + chord progressions	Passive listening	64	–12	–16	MNI	R	100
			–64	–32	–8	MNI	L	

ABX discrimination, judge whether a third stimulus is the same as the first or second stimulus; AX discrimination, same–different judgments on pairs of stimuli; CVs, consonant–vowel syllables; CVCs, consonant–vowel–consonant syllables; L, left; MNI, Montreal Neurological Institute; R, right.

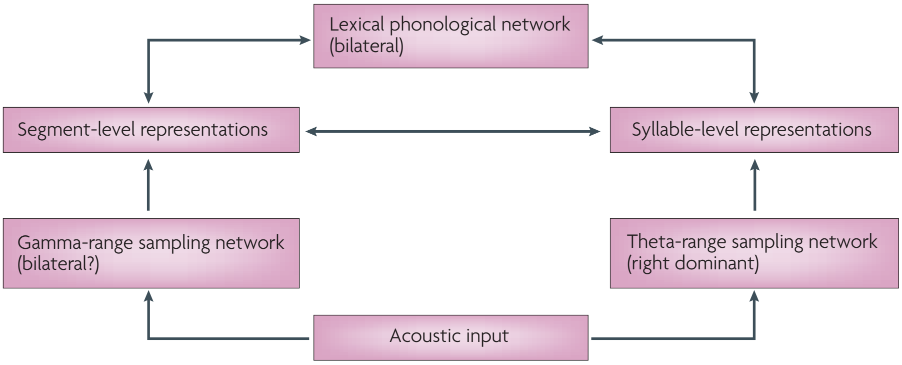
Lexical phonological network
(bilateral)
Segment-level representations
Syllable-level representations
Gamma-range sampling network
(bilateral?)
Theta-range sampling network
(right dominant)
Acoustic input

Figure 2 | **Parallel routes in the mapping from acoustic input to lexical phonological representations.** The figure depicts one characterization of the computational properties of the pathways. One pathway samples acoustic input at a relatively fast rate (gamma range) that is appropriate for resolving segment-level information, and may be instantiated in both hemispheres. The other pathway samples acoustic input at a slower rate (theta range) that is appropriate for resolving syllable level information, and may be more strongly represented in the right hemisphere. Under normal circumstances these pathways interact, both within and between hemispheres, yet each appears capable of separately activating lexical phonological networks. Other hypotheses for the computational properties of the two routes exist.

Phonological processing and the STS. Beyond the earliest stages of speech recognition, there is accumulating evidence and a convergence of opinion that portions of the STS are important for representing and/or processing phonological information[6,16,26,42,51]. The STS is activated by language tasks that require access to phonological information, including both the perception and production of speech[51], and during active maintenance of phonemic information[52,53]. Portions of the STS seem to be relatively selective for acoustic signals that contain phonemic information when compared with complex non-speech signals (FIG. 3a). STS activation can be modulated by the manipulation of psycholinguistic variables that tap phonological networks[54], such as phonological neighbourhood density (the number of words that sound similar to a target word) (FIG. 3b). Thus, a range of studies converge on the STS as a site that is crucial to phonological-level processes. Although many authors consider this system to be strongly left dominant, both lesion and imaging (FIG. 3) evidence suggest a bilateral organization with perhaps a mild leftward bias.

A number of studies have found activation during speech processing in anterior portions of the STS[13,27,29,30], leading to suggestions that these regions have an important and, according to some papers, exclusive role in ventral stream phonological processes[55]. This is in contrast to the typical view that posterior areas form the primary projection targets of the ventral stream[5,6]. However, many of the studies that highlighted anterior regions used sentence- or narrative-level stimuli contrasted against a low-level auditory control. It is therefore impossible to determine which levels of language processing underlie these activations. Furthermore, several recent functional imaging studies have implicated anterior temporal regions in sentence-level processing[56–60], suggesting that syntactic or combinatorial processes might drive much of the anterior temporal activation. In addition, the claim that posterior STS regions are not part of the ventral stream is dubious given the extensive evidence that left posterior temporal lobe disruption leads to auditory comprehension deficits[6,61,62]. It is possible that the ventral projection pathways extend both posteriorly and anteriorly[30]. We suggest that the crucial portion of the STS that is involved in phonological-level processes is bounded anteriorly by the most anterolateral aspect of Heschl's gyrus and posteriorly by the posterior-most extent of the Sylvian fissure. This corresponds to the distribution of activation for 'phonological' processing, depicted in FIG. 3.

Lexical, semantic and grammatical linkages. Much research on speech perception seeks to understand processes that lead to the access of phonological codes. However, during auditory comprehension, the goal of speech processing is to use these codes to access higher-level representations that are vital to comprehension. There is strong evidence that posterior middle temporal regions are involved in accessing lexical and semantic information. However, there is debate about whether other anterior temporal lobe (ATL) regions also participate in lexical and semantic processing, and whether they might contribute to grammatical or compositional aspects of speech processing.

Damage to posterior temporal lobe regions, particularly along the middle temporal gyrus, has long been associated with auditory comprehension deficits[61,63,64], an effect that was recently confirmed in a large-scale study involving 101 patients[61]. Data from direct cortical stimulation studies corroborate the involvement of the middle temporal gyrus in auditory comprehension, but also indicate a much broader network involving most of the superior temporal lobe (including anterior portions), and the inferior frontal lobe[65]. Functional imaging studies have also implicated posterior middle temporal regions in lexical semantic processing[66–68]. These findings do not preclude the involvement of more anterior regions in lexical semantic access, but they do make a strong case for the dominance of posterior regions in these processes. We suggested previously that posterior middle temporal regions supported lexical and semantic access in the form of a sound-to-meaning interface network[5,6]. According to this hypothesis, semantic information is represented in a highly distributed fashion throughout the cortex[69], and middle posterior temporal regions are involved in the mapping between phonological representations in the STS and widely distributed semantic representations. Most of the evidence reviewed above indicates a left-dominant organization for this middle temporal gyrus network. However, the finding that the right hemisphere can comprehend words reasonably well suggests that there is some degree of bilateral capability in lexical and semantic access, but that there are perhaps some differences in the computations that are carried out in each hemisphere.

ATL regions have been implicated in both lexical/semantic and sentence-level processing (syntactic and semantic integration processes). Patients with semantic dementia have atrophy involving the ATL bilaterally, along with deficits on lexical tasks such as naming, semantic association and single-word comprehension[70] , which has been used to argue for a lexical or semantic function for the ATL[29,30]. However, these deficits might be more general, given that the atrophy involves a number of regions in addition to the lateral ATL, including the bilateral inferior and medial temporal lobe, bilateral caudate nucleus and right posterior thalamus, among

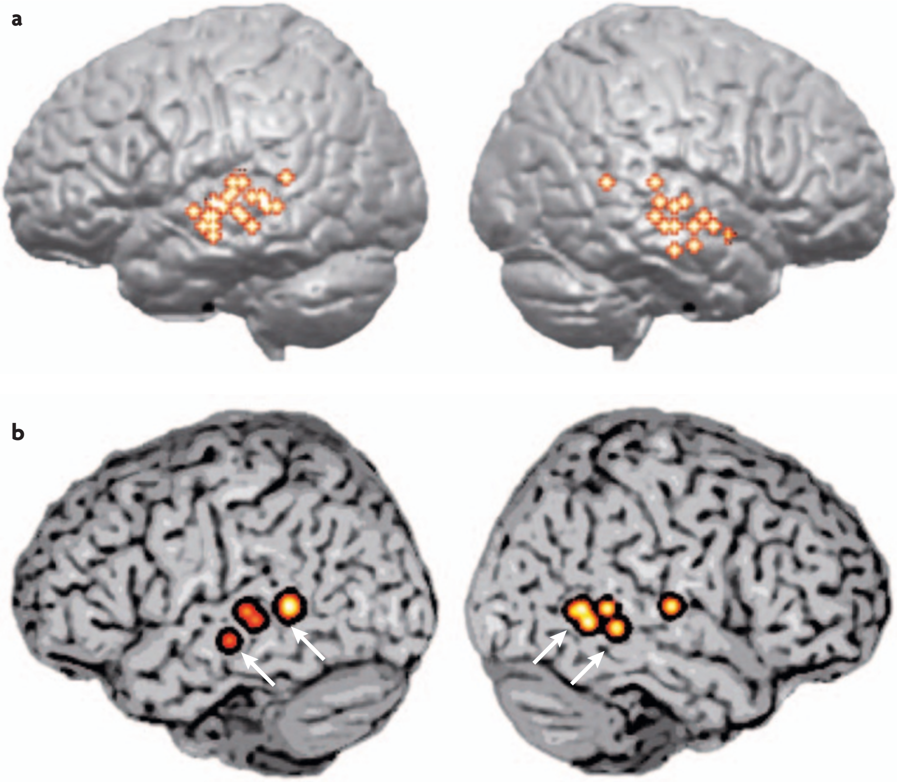

Figure 3 | **Lexical phonological networks in the superior temporal sulcus.** **a** | Distribution of activation foci in seven recent studies of speech processing using sublexical stimuli contrasted with non-speech controls[26,31,107–111]. All coordinates are plotted in Montreal Neurological Institute (MNI) space (coordinates originally reported in Talairach space were transformed to MNI space). Study details and coordinates are presented in TABLE 1. **b** | Highlighted areas represent sites of activation in a functional MRI study contrasting high neighbourhood density words (those with many similar sounding neighbours) with low neighbourhood density words (those with few similar sounding neighbours). The middle to posterior portions of the superior temporal sulcus in both hemispheres (arrows) showed greater activation for high density than low density words (coloured blobs), presumably reflecting the partial activation of larger neural networks when high density words are processed. Neighbourhood density is a property of lexical phonological networks[112]; thus, modulation of neural activity by density manipulation probably highlights such networks in the brain. Reproduced, with permission, from REF. 54 © (2006) Lippincott Williams & Wilkins.

others[70]. This renders the suggested link between lexical deficits and the ATL tenuous.

Higher-level syntactic and compositional semantic processing might involve the ATL. Functional imaging studies have found portions of the ATL to be more active while individuals listen to or read sentences rather than unstructured lists of words or sounds[56–58,60]. This structured-versus-unstructured effect is independent of the semantic content of the stimuli, although semantic manipulations can modulate the ATL response somewhat[60]. Damage to the ATL has also been linked to deficits in comprehending complex syntactic structures[71]. However, data from semantic dementia is contradictory, as these patients are reported to have good sentence-level comprehension[70].

In summary, there is strong evidence that lexical semantic access from auditory input involves the posterior lateral temporal lobe. In terms of syntactic and compositional semantic operations, neuroimaging evidence is converging on the ATL as an important component of the computational network[58–60]; however, the neuropsychological evidence remains equivocal.

Dorsal stream: sound to action

It is generally agreed that the auditory ventral stream supports the perception and recognition of auditory objects such as speech; although, as outlined above, there are a number of outstanding issues concerning the precise mechanisms and regions involved[5,6,18,19,55]. There is less agreement regarding the functional role of the auditory dorsal stream. The earliest proposals argued for a role in spatial hearing, a 'where' function[18], similar to the dorsal 'where' stream proposal in the cortical visual system[17]. More recently we, along with others, have suggested[5,6,22,55] that the auditory dorsal stream supports an interface with the motor system, a proposal that is similar to recent claims that the dorsal visual stream has a sensorimotor integration function[72,73] (BOX 2).

The need for auditory–motor integration. The idea of auditory–motor interaction in speech is not new. Wernicke's classic model of the neural circuitry of language incorporated a direct link between sensory and motor representations of speech, and argued explicitly that sensory systems participated in speech production[1]. Motor theories of speech perception also assume a link between sensory input and motor speech systems[43]. However, the simplest argument for the necessity of auditory–motor interaction in speech comes from development. Learning to speak is essentially a motor learning task. The primary input to this is sensory, speech in particular. So, there must be a neural mechanism that both codes and maintains instances of speech sounds, and can use these sensory traces to guide the tuning of speech gestures so that the sounds are accurately reproduced[1,74].

We have suggested that speech development is a primary and crucial function of the proposed dorsal auditory–motor integration circuit, and that it also continues to function in adults[5,6]. Evidence for the latter includes the disruptive effects of altered auditory feedback on speech production[75,76], articulatory decline in late-onset deafness[77] and the ability to acquire new vocabulary. We also propose that this auditory–motor circuit provides the basic neural mechanisms for phonological short-term memory[53,78,79].

We suggest that there are at least two levels of auditory–motor interaction — one involving speech segments and the other involving sequences of segments. Segmental-level processes would be involved in the acquisition and maintenance of basic articulatory phonetic skills. Auditory–motor processes at the level of sequences of segments would be involved in the acquisition of new vocabulary, and in the online guidance of speech sequences[80]. We propose that auditory–motor interactions in the acquisition of new vocabulary involve generating a sensory representation of the new word that codes the sequence of segments or syllables. This sensory representation can then be used to guide motor articulatory sequences. This might involve true feedforward mechanisms (whereby sensory codes for a speech sequence are translated into a motor speech sequence), feedback monitoring mechanisms, or both. As the word becomes familiar, the nature of the sensory–motor interaction might change. New, low-frequency or more complex words might require incremental motor coding and thus more sensory guidance than known, high-frequency or more simple words, which might become 'automated' as motor chunks that require little sensory guidance. This hypothesis is consistent with a large motor learning literature showing shifts in the mechanisms of motor control as a function of learning[81–83].

Lesion evidence for a sensorimotor dorsal stream. Damage to auditory-related regions in the left hemisphere often results in speech production deficits[63,84], demonstrating that sensory systems participate in motor speech. More specifically, damage to the left dorsal STG or the temporoparietal junction is associated with conduction aphasia, a syndrome that is characterized by good comprehension but frequent phonemic errors in speech production[8,85]. Conduction aphasia has classically been considered to be a disconnection syndrome involving damage to the arcuate fasciculus. However, there is now good evidence that this syndrome results from cortical dysfunction[86,87]. The production deficit is load-sensitive: errors are more likely on longer, lower-frequency words and verbatim repetition of strings of speech with little semantic constraint[85,88]. Functionally, conduction aphasia has been characterized as a deficit in the ability to encode phonological information for production[89].

We have suggested that conduction aphasia represents a disruption of the auditory–motor interface system[6,90], particularly at the segment sequence level. Comprehension of speech is preserved because the lesion does not disrupt ventral stream pathways and/or because right-hemisphere speech systems can compensate for disruption of left-hemisphere speech perception systems. Phonological errors occur because sensory representations of speech are prevented from providing online guidance of speech sound sequencing; this effect is most pronounced for longer, lower-frequency or novel words, because these words rely on sensory involvement to a greater extent than shorter, higher-frequency words, as discussed above. Directly relevant to this claim, a recent functional imaging study showed that activity in the region that is often affected in conduction aphasia is modulated by word length in a covert naming task[91].

Functional imaging evidence for a sensorimotor dorsal stream. Recent functional imaging studies have identified a neural circuit that seems to support auditory–motor interaction[52,53]. Individuals were asked to listen to pseudowords and then subvocally reproduce them. On the assumption that areas involved in integrating sensory and motor processes would have both sensory- and motor-response properties[92,93] (BOX 2), the analyses focused on regions that were active during both the perceptual and motor-related phases of the trial. A network of regions was identified, including the posterior STS bilaterally, a left-dominant site in the Sylvian fissure at the boundary between the parietal and temporal lobes (area Spt), and left posterior frontal regions (FIG. 4)[52,53]. We propose that the posterior STS (bilaterally) supports sensory coding of speech, and that area Spt is involved in translation between those sensory codes and the motor system[53]. This hypothesis is motivated by several considerations: the participation of bilateral areas, including the STS, in sensory and comprehension aspects of speech perception; the left-dominant organization of speech production; the fact that Spt and inferior frontal areas are more tightly correlated in their activation timecourse than STS and inferior frontal areas[52]; and the lesions associated with conduction aphasia that coincide with the location of Spt[63].

Follow-up studies have shed more light on the functional properties of area Spt. Spt activity is not specific to speech: it is activated equally well by the perception and reproduction (via humming) of tonal sequences[53]. Spt was also equally active during the reproduction of sensory stimuli that were perceived aurally (spoken words) or visually (written words)[78], indicating that access to this network is not necessarily restricted to the auditory modality. However, Spt activity is modulated by a manipulation of output modality. Skilled pianists were asked to listen to novel melodies and then to covertly reproduce them by either humming or imagining playing them on a keyboard. Spt was significantly more active during the humming condition than the playing condition, despite the fact that the playing condition is more difficult (G.H., unpublished observations). These results indicate that Spt might be more tightly coupled to a specific motor effector system, the vocal tract, than to a specific sensory system. It might therefore be more accurate to view area Spt as part of a sensorimotor integration circuit for the vocal tract, rather than specifically as part of an auditory–motor integration circuit. This places area Spt within a network of parietal lobe sensorimotor interface regions, each of which seems to be organized primarily around a different motor effector system[72,94], rather than particular sensory systems.

Area Spt is located within the planum temporale (PT), a region that has been at the centre of much recent functional–anatomical debate. Although the PT has traditionally been associated with speech processing, this view has been challenged on the basis of functional imaging studies that find activation in the PT for various acoustic signals (tone sequences, music, spatial signals), as well as some non-acoustic signals (visual speech, sign language, visual motion)[95]. One recent proposal[95] is that the PT functions as a computational hub that takes input from the primary auditory cortex and performs a segregation and spectrotemporal pattern-matching operation; this leads to the output of sound object information, which is processed further in lateral temporal lobe areas, and spatial position information, which is processed further in parietal structures. This proposal is squarely within the framework of a dual-stream model of auditory processing, but differs *prima facie* from our view in terms of the proposed function of both the PT region (computational hub versus sensorimotor integration) and the dorsal stream (spatial versus sensorimotor integration). It is worth noting that the PT is probably not homogeneous. In fact, four different cytoarchitectonic fields have been observed in this region[96]. It is therefore conceivable that one region of the PT serves as a computational hub (or some other function) and another portion computes sensorimotor transformations. Within-subject experiments with multiple stimulus and task conditions are needed to sort out a possible parcellation of the PT. Furthermore, the concept that the auditory dorsal stream supports sensorimotor integration is not incompatible with it also supporting a spatial hearing function (that is, both 'how' and 'where' streams could coexist). Finally, the fact that the PT region receives various inputs, both acoustic and non-acoustic, is perfectly consistent with our proposed model. Speech, tones, music, environmental sounds and visual speech can all be linked to vocal tract gestures (such as speech, humming and naming). Spatial signals, however, should not activate the same sensory–vocal tract network, unless the response task involves speech. It will be instructive to determine whether sensory–vocal tract tasks and

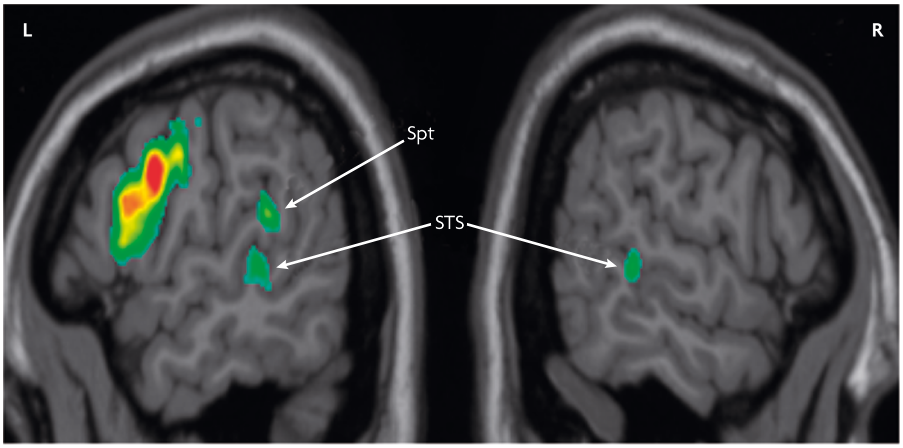

Figure 4 | **Regions activated in an auditory–motor task for speech.** Coloured regions demonstrated sensorimotor responses: they were active during both the perception of speech and the subvocal reproduction of speech. Note the bilateral activation in the superior temporal sulcus (STS), and the unilateral activation of area Spt and frontal regions. Reproduced, with permission, from REF. 113 © (2003) Cambridge Univ. Press.

spatial hearing tasks activate the same PT region. Finally, the dorsal stream circuit that we are proposing is strongly left dominant, and, within the left hemisphere, involves only a portion of the PT. Consequently, our proposal is not intended as a general model of cortical auditory processing or PT function, which leaves plenty of cortical regions for other functional circuits to occupy.

Summary and future perspectives

The dual-stream model that is outlined in this article aims to integrate a wide range of empirical observations, including basic perceptual processes[48], aspects of speech development[6] and speech production[91,97], linguistic and psycholinguistic facts[48,54,98], verbal working memory[5,6,53], task-related effects[5,6], sensorimotor integration circuits[6,53,90], and neuropsychological facts including patterns of sparing and loss in aphasia[6,90]. The basic concept, that the acoustic speech network must interface with conceptual systems on the one hand, and motor–articulatory systems on the other, has proven its utility in accounting for an array of fundamental observations, and fits with similar proposals in the visual[73] — and now somatosensory[99] — domains. This similarity in organization across sensory systems suggests that the present dual-stream proposal for cortical speech processing might be a reflection of a more general principle of sensory system organization.

Assuming that the basic dual-stream framework is correct, the main task for future research will be to specify the details of the within-stream organization and computational operations. We have made some specific hypotheses in this regard; for example, that the ventral stream is itself composed of parallel pathways that could differ in terms of their sampling rate, and that the dorsal stream might also involve parallel systems that differ in scale (segments versus sequences of segments). Further empirical work is required to test these hypotheses, and to develop more explicit models of the architecture and computations involved. Finally, a major challenge will be to understand how these neural models relate to linguistic and psycholinguistic models of language structure and processing.

Gregory Hickok is at the Department of Cognitive Sciences and the Center for Cognitive Neuroscience, University of California, Irvine, California 92697-5100, USA.

David Poeppel is at the Departments of Linguistics and Biology, University of Maryland, College Park, Maryland 20742, USA.

Correspondence to G.H.
e-mail: greg.hickok@uci.edu

doi:10.1038/nrn2113
Published online 13 April 2007

1. Wernicke, C. in *Boston Studies in the Philosophy of Science* (eds Cohen, R. S. & Wartofsky, M. W.) 34–97 (D. Reidel, Dordrecht, 1874/1969).
2. Basso, A., Casati, G. & Vignolo, L. A. Phonemic identification defects in aphasia. *Cortex* **13**, 84–95 (1977).
3. Blumstein, S. E., Baker, E. & Goodglass, H. Phonological factors in auditory comprehension in aphasia. *Neuropsychologia* **15**, 19–30 (1977).
4. Blumstein, S. E., Cooper, W. E., Zurif, E. B. & Caramazza, A. The perception and production of voice-onset time in aphasia. *Neuropsychologia* **15**, 371–383 (1977).
5. Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131–138 (2000).
6. Hickok, G. & Poeppel, D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
7. Miceli, G., Gainotti, G., Caltagirone, C. & Masullo, C. Some aspects of phonological impairment in aphasia. *Brain Lang.* **11**, 159–169 (1980).
8. Damasio, H. & Damasio, A. R. The anatomical basis of conduction aphasia. *Brain* **103**, 337–350 (1980).
9. Caplan, D., Gow, D. & Makris, N. Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology* **45**, 293–298 (1995).
10. Baker, E., Blumsteim, S. E. & Goodglass, H. Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia* **19**, 1–15 (1981).
11. Gainotti, G., Micelli, G., Silveri, M. C. & Villa, G. Some anatomo-clinical aspects of phonemic and semantic comprehension disorders in aphasia. *Acta Neurol. Scand.* **66**, 652–665 (1982).
12. Binder, J. R. *et al.* Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.* **35**, 662–672 (1994).
13. Mazoyer, B. M. *et al.* The cortical representation of speech. *J. Cogn. Neurosci.* **5**, 467–479 (1993).
14. Demonet, J.-F. *et al.* The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768 (1992).
15. Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849 (1992).
16. Price, C. J. *et al.* Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain* **119**, 919–931 (1996).
17. Ungerleider, L. G. & Mishkin, M. in Analysis of visual behavior (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, Massachusetts, 1982).
18. Rauschecker, J. P. Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* **8**, 516–521 (1998).
19. Scott, S. K. Auditory processing — speech, space and auditory objects. *Curr. Opin. Neurobiol.* **15**, 197–201 (2005).
20. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
21. Warren, J. E., Wise, R. J. & Warren, J. D. Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends Neurosci.* **28**, 636–643 (2005).
22. Wise, R. J. S. *et al.* Separate neural sub-systems within 'Wernicke's area'. *Brain* **124**, 83–95 (2001).
23. McClelland, J. L. & Elman, J. L. The TRACE model of speech perception. *Cognit. Psychol.* **18**, 1–86 (1986).
24. Marslen-Wilson, W. D. Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102 (1987).
25. Luce, P. A. & Pisoni, D. B. Recognizing spoken words: the neighborhood activation model. *Ear Hear*. **19**, 1–36 (1998).
26. Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T. & Medler, D. A. Neural substrates of phonemic perception. *Cereb. Cortex* **15**, 1621–1631 (2005).
27. Narain, C. *et al.* Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* **13**, 1362–1368 (2003).
28. Obleser, J., Zimmermann, J., Van Meter, J. & Rauschecker, J. P. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex* 5 Dec 2006 (doi:10.1093/cercor/bhl133).
29. Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. S. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406 (2000).
30. Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. Converging language streams in the human temporal lobe. *J. Neurosci.* **26**, 7328–7336 (2006).
31. Vouloumanos, A., Kiehl, K. A., Werker, J. F. & Liddle, P. F. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* **13**, 994–1005 (2001).
32. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. Speech recognition with primarily temporal cues. *Science* **270**, 303–304 (1995).
33. Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D. Speech perception without traditional speech cues. *Science* **212**, 947–950 (1981).
34. Saberi, K. & Perrott, D. R. Cognitive restoration of reversed speech. *Nature* **398**, 760 (1999).
35. Drullman, R., Festen, J. M. & Plomp, R. Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* **95**, 2670–2680 (1994).
36. Drullman, R., Festen, J. M. & Plomp, R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **95**, 1053–1064 (1994).
37. Zaidel, E. in *The Dual Brain: Hemispheric Specialization in Humans* (eds Benson, D. F. & Zaidel, E.) 205–231 (Guilford, New York, 1985).

38. McGlone, J. Speech comprehension after unilateral injection of sodium amytal. *Brain Lang.* **22**, 150–157 (1984).
39. Poeppel, D. Pure word deafness and the bilateral processing of the speech code. *Cogn. Sci.* **25**, 679–693 (2001).
40. Stefanatos, G. A., Gershkoff, A. & Madigan, S. On pure word deafness, temporal processing, and the left hemisphere. *J. Int. Neuropsychol. Soc.* **11**, 456–470 (2005).
41. Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D. & Schwartz, M. Word deafness: One hundred years later. *J. Neurol. Neurosurg. Psychiatr.* **49**, 489–499 (1986).
42. Binder, J. R. *et al.* Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528 (2000).
43. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
44. Stevens, K. N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* **111**, 1872–1891 (2002).
45. Dupoux, E. in *Cognitive Models of Speech Processing* (eds Altmann, G. & Shillcock, R.) 81–114 (Erlbaum, Hillsdale, New Jersey, 1993).
46. Greenberg, S. & Arai, T. What are the essential cues for understanding spoken language? *IEICE Trans. Inf. Syst.* E87-D, 1059–1070 (2004).
47. Boemio, A., Fromm, S., Braun, A. & Poeppel, D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neurosci.* **8**, 389–395 (2005).
48. Poeppel, D., Idsardi, W. & van Wassenhove, V. Speech perception at the interface of neurobiology and linguistics. *Phil. Trans. Roy. Soc.* (in the press).
49. Zatorre, R. J., Belin, P. & Penhune, V. B. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46 (2002).
50. Schonwiesner, M., Rubsamen, R. & von Cramon, D. Y. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* **22**, 1521–1528 (2005).
51. Indefrey, P. & Levelt, W. J. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
52. Buchsbaum, B., Hickok, G. & Humphries, C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* **25**, 663–678 (2001).
53. Hickok, G., Buchsbaum, B., Humphries, C. & Muftuler, T. Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* **15**, 673–682 (2003).
54. Okada, K. & Hickok, G. Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *Neuroreport* **17**, 1293–1296 (2006).
55. Scott, S. K. & Wise, R. J. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* **92**, 13–45 (2004).
56. Friederici, A. D., Meyer, M. & von Cramon, D. Y. Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang.* **74**, 289–300 (2000).
57. Humphries, C., Willard, K., Buchsbaum, B. & Hickok, G. Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport* **12**, 1749–1752 (2001).
58. Humphries, C., Love, T., Swinney, D. & Hickok, G. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum. Brain Mapp.* **26**, 128–138 (2005).
59. Humphries, C., Binder, J. R., Medler, D. A. & Liebenthal, E. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* **18**, 665–679 (2006).
60. Vandenberghe, R., Nobre, A. C. & Price, C. J. The response of left temporal cortex to sentences. *J. Cogn. Neurosci.* **14**, 550–560 (2002).
61. Bates, E. *et al.* Voxel-based lesion-symptom mapping. *Nature Neurosci.* **6**, 448–450 (2003).
62. Boatman, D. Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* **92**, 47–65 (2004).
63. Damasio, H. in *Acquired aphasia* (ed. Sarno, M.) 45–71 (Academic, San Diego, 1991).
64. Dronkers, N. F., Redfern, B. B. & Knight, R. T. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 949–958 (MIT Press, Cambridge, Massachusetts, 2000).
65. Miglioretti, D. L. & Boatman, D. Modeling variability in cortical representations of human complex sound perception. *Exp. Brain Res.* **153**, 382–387 (2003).
66. Binder, J. R. *et al.* Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* **17**, 353–362 (1997).
67. Rissman, J., Eliassen, J. C. & Blumstein, S. E. An event-related FMRI investigation of implicit semantic priming. *J. Cogn. Neurosci.* **15**, 1160–1175 (2003).
68. Rodd, J. M., Davis, M. H. & Johnsrude, I. S. The neural mechanisms of speech comprehension: fMRI studeis of semantic ambiguity. *Cereb. Cortex* **15**, 1261–1269 (2005).
69. Damasio, A. R. & Damasio, H. in Large-scale neuronal theories of the brain (eds Koch, C. & Davis, J. L.) 61–74 (MIT Press, Cambridge, Massachusetts, 1994).
70. Gorno-Tempini, M. L. *et al.* Cognition and anatomy in three variants of primary progressive aphasia. *Ann. Neurol.* **55**, 335–346 (2004).
71. Dronkers, N. F., Wilkins, D. P., Van Valin, R. D. Jr, Redfern, B. B. & Jaeger, J. J. in *The New Functional Anatomy of Language: A special issue of Cognition* (eds Hickok, G. & Poeppel, D.) 145–177 (Elsevier Science, 2004).
72. Andersen, R. Multimodal integration for the representation of space in the posterior parietal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **352**, 1421–1428 (1997).
73. Milner, A. D. & Goodale, M. A. The visual brain in action (Oxford Univ. Press, Oxford, 1995).
74. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Ann. Rev. Neurosci.* **22**, 567–631 (1999).
75. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
76. Yates, A. J. Delayed auditory feedback. *Psychol. Bull.* **60**, 213–251 (1963).
77. Waldstein, R. S. Effects of postlingual deafness on speech production: implications for the role of auditory feedback. *J. Acoust. Soc. Am.* **88**, 2099–2144 (1989).
78. Buchsbaum, B. R., Olsen, R. K., Koch, P. & Berman, K. F. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* **48**, 687–697 (2005).
79. Jacquemot, C. & Scott, S. K. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* **10**, 480–486 (2006).
80. Bohland, J. W. & Guenther, F. H. An fMRI investigation of syllable sequence production. *Neuroimage* **32**, 821–841 (2006).
81. Doyon, J., Penhune, V. & Ungerleider, L. G. Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia* **41**, 252–262 (2003).
82. Lindemann, P. G. & Wright, C. E. in *Methods, Models, and Conceptual Issues: An Invitation to Cognitive Science.* Vol. 4 (eds Scarborough, D. & Sternberg, E.) 523–584 (MIT Press, Cambridge, Massachusetts, 1998).
83. Schmidt, R. A. A schema theory of discrete motor skill learning. *Psychol. Rev.* **82**, 225–260 (1975).
84. Damasio, A. R. Aphasia. *N. Engl. J. Med.* **326**, 531–539 (1992).
85. Goodglass, H. in *Conduction Aphasia* (ed. Kohn, S. E.) 39–49 (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1992).
86. Anderson, J. M. *et al.* Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke–Geschwind model. *Brain Lang.* **70**, 1–12 (1999).
87. Hickok, G. *et al.* A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci. Lett.* **287**, 156–160 (2000).
88. Goodglass, H. *Understanding Aphasia* (Academic, San Diego, 1993).
89. Wilshire, C. E. & McCarthy, R. A. Experimental investigations of an impairement in phonological encoding. *Cogn. Neuropsychol.* **13**, 1059–1098 (1996).
90. Hickok, G. in *Language and the Brain* (eds Grodzinsky, Y., Shapiro, L. & Swinney, D.) 87–104 (Academic, San Diego, 2000).
91. Okada, K., Smith, K. R., Humphries, C. & Hickok, G. Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport* **14**, 2323–2326 (2003).
92. Murata, A., Gallese, V., Kaseda, M. & Sakata, H. Parietal neurons related to memory-guided hand manipulation. *J. Neurophysiol.* **75**, 2180–2186 (1996).
93. Rizzolatti, G., Fogassi, L. & Gallese, V. Parietal cortex: from sight to action. *Curr. Opin. Neurobiol.* **7**, 562–567 (1997).
94. Colby, C. L. & Goldberg, M. E. Space and attention in parietal cortex. *Ann. Rev. Neurosci.* **22**, 319–349 (1999).
95. Griffiths, T. D. & Warren, J. D. The planum temporale as a computational hub. *Trends Neurosci.* **25**, 348–353 (2002).
96. Galaburda, A. & Sanides, F. Cytoarchitectonic organization of the human auditory cortex. *J. Comp. Neurol.* **190**, 597–610 (1980).
97. Okada, K. & Hickok, G. An fMRI study investigating posterior auditory cortex activation in speech perception and production: evidence of shared neural substrates in superior temporal lobe. *Soc. Neurosci. Abstr.* 288.11 (2003).
98. Hickok, G. Functional anatomy of speech perception and speech production: Psycholinguistic implications. *J. Psycholinguist. Res.* **30**, 225–234 (2001).
99. Dijkerman, H. C. & de Haan, E. H. F. Somatosensory processes subserving perception and action. *Behav Brain Sci.* (in the press).
100. Perenin, M.-T. & Vighetto, A. Optic ataxia: A specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain* **111**, 643–674 (1988).
101. Goodale, M. A., Milner, A. D., Jakobson, L. S. & Carey, D. P. A neurological dissociation between perceiving objects and grasping them. *Nature* **349**, 154–156 (1991).
102. Darwin, C. J. in *Attention and Performance X: Control of Language Processes* (eds Bouma, J. & Bouwhuis, D. G.) 197–209 (Lawrence Erlbaum Associates, London,1984).
103. Diehl, R. L. & Kluender, K. R. On the objects of speech perception. *Ecol. Psychol.* **1**, 121–144 (1989).
104. Liberman, A. M. & Mattingly, I. G. A specialization for speech perception. *Science* **243**, 489–494 (1989).
105. Price, C., Thierry, G. & Griffiths, T. Speech-specific auditory processing: where is it? *Trends Cogn. Sci.* **9**, 271–276 (2005).
106. Whalen, D. H. *et al.* Differentiation of speech and nonspeech processing within primary auditory cortex. *J. Acoust. Soc. Am.* **119**, 575–581 (2006).
107. Benson, R. R., Richardson, M., Whalen, D. H. & Lai, S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage* **31**, 342–353 (2006).
108. Dehaene-Lambertz, G. *et al.* Neural correlates of switching from auditory to speech perception. *Neuroimage* **24**, 21–33 (2005).
109. Jancke, L., Wustenberg, T., Scheich, H. & Heinze, H. J. Phonetic perception and the temporal cortex. *Neuroimage* **15**, 733–746 (2002).
110. Joanisse, M. F. & Gati, J. S. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage* **19**, 64–79 (2003).
111. Rimol, L. M., Specht, K., Weis, S., Savoy, R. & Hugdahl, K. Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage* **26**, 1059–1067 (2005).
112. Vitevitch, M. S. & Luce, P. A. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* **40**, 374–408 (1999).
113. Hickok, G. & Buchsbaum, B. Temporal lobe speech perception systems are part of the verbal working memory circuit: evidence from two recent fMRI studies. *Behav. Brain Sci.* **26**, 740–741 (2003).

Acknowledgements

Supported by grants from the National Institutes of Health.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Hickok's laboratory: http://lcbr.ss.uci.edu
Poeppel's laboratory: http://www.ling.umd.edu/cnl
Access to this interactive links box is free online.