CHAPTER 1

A review of the role of speech motor areas in speech perception

1.1 Introduction

The earliest stages of speech perception in the cortex involve primary auditory cortex in Heschl's gyrus, and surrounding belt and parabelt regions in the superior temporal gyrus; the importance of this region for speech perception has been known for over a century (Wernicke, 1874; Scott & Wise, 2004; Hickok & Poeppel, 2004). However many researchers have been intrigued by the possibility that speech motor regions might also play a role in perception. Alajouanine et al. (1964) reported that phonemic discrimination errors were more pronounced in patients whose speech output contained phonemic jargon than those who produced semantic jargon. Luria (1966) observed that speech perception was compromised as a side effect of articulatory disorders, and proposed that silent articulation was necessary to process speech sounds. Meanwhile phoneticians arrived at similar ideas. The *motor theory of speech perception* (Liberman et al., 1967; Liberman & Mattingly, 1985) proposed that the objects of speech perception are the intended phonetic gestures of the speaker, implying a role for the motor system in the process. Likewise the *direct realist theory* of speech perception stressed the perception of distal objects (articulatory gestures) over the proximal percept of the acoustic signal (Fowler & Rosenblum, 1991).

Whereas the aphasiologists made their claims based on empirical grounds, Liberman and colleagues' argument was primarily theoretical. They observed that the acoustic cues for each phoneme were highly contextually dependent, and reasoned that only at the level of motor control structures would phonetic representations be invariant. Furthermore, they stressed the notion that *parity* must be maintained between acoustic and articulatory representations. Empirical arguments in support of the motor theory were advanced, but ultimately did not prove to be compelling. For instance, categorical perception refers to the fact that stop consonants in particular can only be perceived categorically: listeners are insensitive to changes in voice onset time (VOT) or formant frequency unless they cross a categorical boundary between two phonemes (Liberman et al., 1957). Likewise these sounds are also produced categorically: it is impossible to produce a sound which is halfway between a [p] and a [t]. These data could be explained if perception was in articulatory terms. However categorical perception was demonstrated first in infants before they had begun to speak (Eimas et al., 1971), and then in the chinchilla, a species which will never speak (Kuhl & Miller, 1975). These findings demonstrated that categorical perception is not dependent on articulation, and must be an auditory phenomenon. Moreover electromyographic studies revealed that motor commands reaching muscles are also highly variable subject to coarticulatory demands (see MacNeilage, 1972, for review), suggesting that articulatory representations are not necessarily any more invariant than acoustic ones.

However the idea that the motor system might be involved in speech perception was revitalized with the discovery of “mirror neurons”: premotor neurons which respond to perception as well as production of actions (for review see Rizzolatti et al., 2002; Rizzolatti & Craighero, 2004). Such neurons were first discovered in premotor area F5 of the macaque (di Pellegrino et al., 1992; Gallese et al., 1996), and were subsequently inferred to exist in humans based on TMS (Fadiga et al., 1995; Strafella & Paus, 2000) and neuroimaging (Rizzolatti et al., 1996; Iacoboni et al., 1999) studies. Premotor responses not only to visual actions but also to the sounds of actions have been demonstrated in the macaque (Kohler et al., 2002) and in humans (Aziz-Zadeh et al., 2004). The appeal of mirror neurons is that such a mechanism, by relating action and perception, might hold the key to the maintenance of parity between articulatory and acoustic representations. Theoretical arguments and empirical data have been presented supporting a role for mirror neurons in language (Rizzolatti & Arbib, 1998; Fadiga et al., 2002; Watkins et al., 2003; Wilson et al., 2004; Pulvermüller et al., 2006).

Researchers interested in the neural basis of speech perception have also hypothesized a role for speech motor areas in the process. Hickok & Poeppel (2000, 2004) proposed that there are two streams of processing involved in speech perception: a ventral stream involving the superior temporal sulcus (STS), middle temporal gyrus (MTG) and inferior temporal gyrus (ITG) where sound-based representations interface with lexical conceptual representations, and a dorsal auditory-motor stream (see also Wise et al., 2001; Scott & Johnsrude, 2003; Scott & Wise, 2004; Warren et al., 2005 for related proposals). The auditory-motor pathway performs bidirectional mapping between auditory and motor representations of speech, and is responsible for the maintenance of parity (Liberman & Mattingly, 1985). Thus it instantiates *internal models* (Haruno et al., 2001) of the relationship between articulatory and acoustic forms. The dorsal stream is proposed to project to frontal cortex via a region on the posterior superior temporal plane (Buchsbaum et al., 2001; Wise et al., 2001; Hickok et al., 2003). The arcuate fasciculus, which has traditionally been held to link posterior and anterior perisylvian language areas (Geschwind, 1965), may be an anatomical substrate of the dorsal stream. In the macaque the dorsal auditory pathway projects to prefrontal (Deacon, 1992; Romanski et al., 1999; Petrides, 2005) and premotor (Deacon, 1992) cortex. Connections may also be made via inferior parietal cortex (Hickok & Poeppel, 2000). The pathway for mapping between acoustic and articulatory representations is hypothesized to be important for both production and perception of speech. In production, speech gestures would be planned in acoustic space and then mapped onto articulatory outputs (Guenther et al., 1998). In perception, Hickok & Poeppel (2000, 2004) propose that the auditory-motor pathway is important mainly for explicit access to phonetic segments such as when performing phonological or phonetic tasks, or when repeating speech. Another proposal is that the auditory-motor pathway is particularly important in situations where the acoustic signal is degraded, such as in noise, or when processing a second language (Callan et al., 2004).

In this review chapter we discuss neuroscientific evidence for the role of motor and premotor regions in speech perception. We begin with a brief discussion of the functional neuroanatomy of speech motor areas, in order to provide a point of reference. Then we discuss relevant studies from a range of methodologies: functional neuroimaging, single pulse and repetitive transcranial magnetic stimulation (TMS), lesion studies, and direct electrical stimulation of the cortex. Two issues frequently arise. Firstly, it is often not clear whether motor areas are involved in speech perception specifically or other cognitive processes. Secondly, the relationship between phonological and phonetic tasks and real world language comprehension is not necessarily straightforward.

In the following three chapters of the dissertation, three fMRI studies are presented which support a role for speech motor areas in speech perception. In the first study (chapter 2), subjects listened passively to monosyllables, and produced the same speech sounds. A region of overlap between the perception and production tasks was found in premotor cortex. Chapter 3 describes a study examining neural responses to unfamiliar non-native phonemes varying in the extent to which they can be successfully articulated. Precentral areas were activated by passive speech perception, functionally connected to superior temporal cortex, and distinguished native from non-native phonemes. In superior temporal regions, activity covaried with producibility of non-native phonemes, suggesting that candidate phonemic categorizations generated in premotor cortex may be compared to the acoustic input in temporal cortex. In the third study, described in chapter 4, subjects were presented with auditory or audiovisual narratives, and model-free intersubject correlational analyses were employed to reveal areas that were modulated in a consistent way across subjects during narrative comprehension. The intersubject correlational analyses revealed an extended network of areas typically not reported in previous studies of narrative speech comprehension, including extensive bilateral inferior frontal and premotor regions. Chapter 5 summarizes and concludes the dissertation. Note that the three studies which comprise the empirical body of the dissertation (chapter 2: Wilson et al., 2004; chapter 3: Wilson & Iacoboni, 2006; chapter 4: Wilson et al., submitted) are all discussed in advance in this review chapter.

1.2 Functional neuroanatomy of frontal speech motor areas

A selection of representative neuroimaging studies of speech production are summarized in Table 1.1 and coordinates falling in lateral frontal areas or the insula are plotted in Figure 1.1. All coordinates have been plotted relative to probabilistic cytoarchitectonic maps (Geyer et al., 1996; Amunts et al., 1999; Geyer, 2004), and as we discuss activation peaks in this review we will state their locations with reference to these maps, rather than using any anatomical designations which may have been provided in the original references. There is evidence for the involvement of four lateral frontal regions in speech production: primary motor cortex, ventral premotor cortex, Brodmann area (BA) 44 in the posterior inferior frontal gyrus, and dorsolateral prefrontal cortex (DLPFC), as well as the anterior insula. Other areas known to be involved in speech production but outside of the scope of this review include the supplementary motor area (Penfield, 1950), the posterior superior temporal gyrus (Hickok & Poeppel, 2004; Scott & Wise, 2004), the cerebellum, and subcortical structures.

The primary motor cortex (M1) is composed of two distinct cytoarchitectonic regions denoted BA 4a and 4p (Geyer et al., 1996). Speech production activations cluster very tightly in these regions, particularly in the left hemisphere. Several meta-analyses of speech production of largely non-overlapping datasets have also reported peaks in this

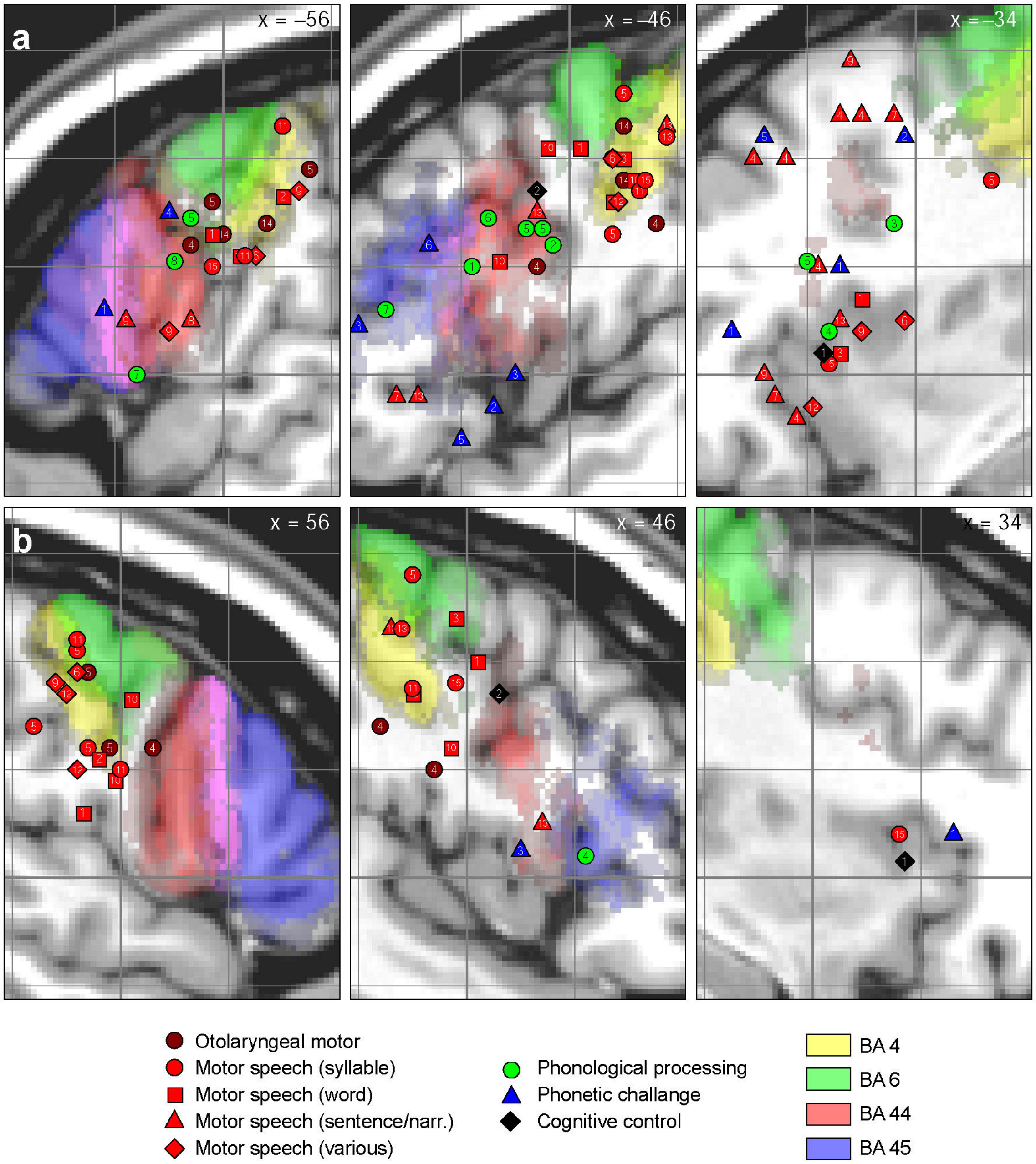

Figure 1.1 Activation peaks from studies of speech production and orolaryngeal motor control, phonological processing, and speech comprehension under phonetically challenging conditions, in left (a) and right (b) lateral frontal regions and the insula. Three sagittal slices are shown for each hemisphere. Points were plotted on the *x* = ±56 slice if they fell in the range [±51, ±70], on the *x* = ±46 slice if they fell in the range [±41, ±50], or on the *x* = ±34 slice if they fell in the range [±20, ±40]. The studies included in this meta-analysis are reported in Tables 1.1, 1.2 and 1.3. The

speech production and phonological processing studies depicted are representative, not exhaustive. The numbers on the symbols relate each peak to the study in which it was reported. Also included are peaks from a review of the role of the inferior frontal junction in cognitive control (Brass et al., 2005: black diamond 1), and regions associated with decision making in phonetic discrimination (Binder et al., 2004: black diamond 2). The background image is a publicly available T1 weighted image which is the average of 27 scans of a single individual (Holmes et al., 1998). This image was registered to MNI space (Mazziotta et al., 2001) based on the procedure described by Collins et al. (1994). Probabalistic cytoarchitectonic maps of Brodmann Areas (BA) 4 (Geyer et al., 1996), BA 6 (Geyer, 2004), BA 44 (Amunts et al., 1999) and BA 45 (Amunts et al., 1999) are shown with shaded color. These maps are publicly available in the same space as the single subject background image (Eickhoff et al., 2005); post-mortem brains were normalized to this template by means of a high-dimensional elastic warping algorithm (Mohlberg et al., 2003; Amunts et al., 2004). Note that the publicly available probabilistic cytoarchitectonic maps were further translated by a few millimeters in the *y* and *z* planes in order to have to anterior commisure (AC) fall at (0, 0, 0) (Eickhoff et al., 2005); we inverted this translation to return the maps to standard MNI space in which the AC is not quite at (0, 0, 0). The maps for BA 4a and BA 4p were summed together to obtain a composite map. Maximum opacity of 40% corresponds to presence of the area in 50% or more of subjects; where the area was found in less than 50% of subjects, colors are increasingly faded. Voxels which belonged to one area in some subjects and another in other subjects are shown with blended colors. These images, and those in the following three figures, were created with a custom MATLAB program.

region (Fox et al., 2001; Turkeltaub et al., 2002; Brown et al., 2005). One study has argued that in individual subject analyses, separate peaks can be observed in areas 4a and 4p (Wilson et al., 2004), but the courseness of neurovascular coupling could also be a significant factor at the single subject level, so a functional distinction between these two segments remains to be confirmed.

Ventral and anterior to orofacial primary motor cortex lies a ventral premotor region (PMv) which also contains orofacial representations (Rizzolatti et al., 2002). Several

Table 1.1 Speech production and orolaryngeal motor studies

#	Study	Method	*N*	Task	Control	Frontal peaks		
						x	*y*	*z*
1	Petersen et al. (1988)	PET*	17	Repeat auditory words	Listen to auditory words	–46	–14	42
						45	–10	40
						–34	–2	14
						62	–19	12
						–52	–10	26
2	Paus et al. (1993)	PET*	8	Hear words and say preassociated words	Fixation	–52	–13	35
						63	–5	24
3	Bookheimer et al. (1995)	PET*	16	Name objects	View visual non-object stimuli	–34	14	5
						–42	–12	43
						42	–5	52
4	Braun et al. (1997)	PET	20	Orolaryngeal motor task[a]	Rest	–52	6	24
						–46	6	20
						–48	–16	28
						52	6	24
						48	–6	20
						44	–16	28
				Narratives and sentence construction	Orolaryngeal motor task	–24	24	40
						–38	22	–8
						–38	18	20
						–26	10	48
				Paced and rote speech	Orolaryngeal motor task	–20	30	40
						–26	14	48
5	Lotze et al. (2000)	fMRI		Lip pursing[a]	Rest	–52	–16	38
						54	–6	38
				Move tip of tongue[a]	Rest	–52	2	32
						66	–2	24
				Say [pa]	Rest	–48	–10	52
						48	–10	56
				Say [ta]	Rest	–46	–8	26
						60	–6	24
				Say [ka]	Rest	–36	–14	36
						64	–16	28
				Say [pataka]	Rest	–44	–8	40
						54	–8	42
6	Fox et al. (2001)[b]	PET*	30	Various tasks	Fixation	–46	–10	43
						53	–10	41
						–55	–7	24
						–30	2	11

7	Braun et al. (2001)	PET	12	Narratives	Orolaryngeal motor task	–44	32	–4
						–40	4	48
						–30	26	–4
8	Indefrey et al. (2001)	PET	12	Sentence scene description	Single word scene description	–54	6	10
9	Blank et al. (2002)	PET	8	Propositional speech, nursery rhyme, and counting	Rest	–52	–14	34
						52	–12	36
						–60	10	8
						–36	10	8
				Nursery rhyme	Counting	–52	18	10
						–36	28	0
				Propositional speech	Nursery rhyme and counting	–36	12	58
10	Turkeltaub et al. (2002)	PET	N/A[c]	Single word reading	Various	–48	–12	36
						44	–10	34
						–42	13	21
						–50	4	42
						55	2	33
						–51	–3	22
						53	–1	18
						–46	–8	32
						48	–3	24
11	Wilson et al. (2004)	fMRI	10	Produce syllables	Rest	–51	–11	46
						–45	–13	34
						–56	–4	22
						56	–8	44
						48	–10	35
						60	0	20
12	Brown et al. (2005)	PET and fMRI*	N/A[d]	Various	Various	–49	–11	34
						–36	20	–6
						55	–12	36
						57	–9	21
13	Kemeny et al. (2005)	fMRI[e]	6	Sentence construction	Rest	42	14	10
						–40	14	10
						–46	6	30
						–48	28	–4
						50	–14	46
						–48	–18	46
				Syllable generation	Rest	50	–12	46
						–48	–18	44

14	Pulvermüller et al. (2006)	fMRI		Lip movement[a]	Rest	–48	–10	36
						–50	–10	46
				Tongue movement[a]	Rest	–56	–8	28
						–56	0	26
15	Wilson & Iacoboni (2006)[f]	fMRI	12	Syllable production	Rest	–60	2	20
						–50	–14	36
						48	–2	36
						–30	16	2
						34	16	8

Note. Asterisks in this and subsequent tables indicates that coordinates were transformed from Talairach to MNI space using the MATLAB program tal2mni, by M. Brett, downloaded from: http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach.

[a] These are mouth motor, not speech production conditions.

[b] Fox et al. (2001) also performed a meta-analysis, however since we have included those studies here, we report only the original data from Fox et al. (2001).

[c] This is a meta-analysis of 11 studies.

[d] This is a meta-analysis of 8 studies.

[e] Blood oxygen level dependent (BOLD) and arterial spin labeling (ASL) fMRI were performed; here we include only the ASL results since the authors concluded that these were more reliable.

[f] Speech production peaks were not reported in this study; we obtained them from our original data.

imaging studies have reported peaks of activation in the left PMv (Petersen et al., 1988; Fox et al., 2001; Turkeltaub et al., 2002; Wilson et al. 2004; Wilson & Iacoboni, 2006). Most authors consider this to be a ventral sector of BA 6 (e.g. Brodmann, 1909; Sarkissov et al., 1955; Rizzolatti et al., 2002; Petrides, 2005). However recent detailed cytoarchitectonic studies suggest that the area on the frontal part of the subcentral gyrus is distinct from BA 6, and from any other area labeled by Brodmann (Geyer, 2003; Eickhoff et al., 2006b). This region is granular (S. Eickhoff, personal communication, August 2, 2006) but also contains large pyramidal cells in layer V (Eickhoff et al., 2006b), and it may be involved with visceral sensorimotor processes (Eickhoff et al., 2006a) which could include speech-related musculature such as the larynx and diaphragm. In the cytoarchitectonic maps used as reference points in this review, the region in question falls ventral to BA 6 and is not labeled. We will refer to this region as PMv and remain agnostic regarding its cytoarchitectonic characterization.

Other imaging studies have reported more anterior activation peaks that likely fall in left BA 44 (Indefrey et al., 2001; Blank et al., 2002; Turkeltaub et al., 2002; Kemeny et al., 2005). It is notable that most of the studies reporting peaks in BA 44 did so for contrasts which highlighted higher-level syntactic and semantic processes: scene description in sentences versus scene description with unconnected words (Indefrey et al., 2001); propositional speech versus nursery rhymes and counting (Blank et al., 2002) and sentence generation but not syllable generation (Kemeny et al., 2005).

Most authors consider Broca’s area to be comprised of BAs 44 and 45 in the left hemisphere, or the pars opercularis and pars triangularis of the inferior frontal gyrus. However the correspondence between these two cytoarchitectonic regions and the gyral anatomy is weak (Amunts et al., 1999). Furthermore, there are significant differences between the cytoarchitecture of areas 44 and 45: for instance, BA 44 is dysgranular whereas BA 45 is granular. We concur with a recent argument that there is presently no motivation for a superordinate structure (“Broca’s area”) comprising these two regions (Hagoort, 2005). No speech production studies reported activation peaks in BA 45. It is important to note though that the opposite conclusion was reached by Horwitz et al. (2003) who reanalyzed PET speech production data from two studies included in the present sample (Braun et al., 1997; Braun et al. 2001) with respect to the same probabilistic cytoarchitectonic maps employed here. The authors concluded that BA 45 was activated by higher-level speech production processes (and also sign language production), whereas BA 44 was more involved in the generation of complex articulatory movements. The methods used to combine imaging and cytoarchitectonic maps were quite different to the present method of plotting peaks, however it is unclear precisely which factor(s) are responsible for the different results. Based on the larger sample of speech production studies considered here, the weight of the evidence favors BA 44 over BA 45 as the inferior frontal region most consistently involved in higher-level speech production processes.

What is clear is that more anterior areas are implicated in higher-level aspects of speech production, whereas more posterior areas are concerned with articulatory processes. Data from cortical stimulation support this basic distinction. Penfield and Roberts (1959) report that speech arrest can be obtained by stimulating a range of points in the left hemisphere. These points span an area from the posterior IFG to the central sulcus, and so presumably include BA 44, PMv and M1. The more posterior points interfere directly with motor control of the speech organs and are accompanied by orofacial movements, whereas the more anterior points do not produce orofacial movements but apparently interfere with a higher level of speech planning. Similar results have been reported using repetitive transcranial magnetic stimulation (rTMS) to produce speech arrest (Stewart et al., 2001). In that study, stimulation of more posterior, dorsal sites produced speech arrest with concomitant activation of the mentalis muscle, whereas stimulation of more anterior, ventral sites produced pure speech arrest. It is unclear whether the posterior site of Stewart et al. (2001) corresponds to PMv or M1.

In the right hemisphere, activations are reliably observed in primary motor cortex, and meta-analyses have reported peaks in right BA 4 (Turkeltaub et al., 2002; Brown et al., 2005). Speech arrest in the right hemisphere is found only from more posterior sites which interfere with motor control directly (Penfield & Roberts, 1959; Stewart et al., 2001). Activations in right BA 44 are rarely observed, but surprisingly a number of studies have reported activations in right PMv (Petersen et al., 1988; Turkeltaub et al., 2002; Wilson et al., 2004); even one meta-analysis found reliable activation in right PMv (Brown et al., 2005). These findings suggest that PMv patterns with M1 in that it is bilaterally involved in speech production. It is likely that right hemisphere stimulations that produce speech arrest probably include stimulations to PMv as well as M1. Some right hemisphere activations in areas besides M1 and PMv may reflect atypical language dominance in subsets of subjects; even though left-handed subjects are excluded from most studies, a minority of right-handers also have atypical language dominance (Knecht et al., 2003).

Two further speech production regions appear in Figure 1.1 but are rarely if ever implicated in speech perception. DLPFC was activated only in studies requiring the generation of narratives (Braun et al., 1997; Braun et al., 2001; Blank et al., 2002), and probably plays a high-level role in speech planning; lesions to this area typically produce transcortical motor aphasia. The anterior insula was activated in most studies regardless of the level of language produced (see Table 1.1 for references), and is believed to be a crucial region for coordinating musculature for speech (Dronkers, 1996; Ackermann & Riecker, 2004). Frontal lesions typically do not produce a permanent speech deficit unless they extend to the insula (Mohr, 1976).

1.3 Phonological and phonetic tasks

One of the earliest PET studies of language showed that a phonological task—detecting whether the final consonants of a pair of words or nonwords matched or not—activated a region of the left IFG (Zatorre et al., 1992). The baseline for this task was passive listening to the pairs of syllables. A number of other studies have also reported left IFG activations for phonological processing tasks (Figure 1.1, Table 1.2; for review see Poldrack et al., 1999; Burton, 2001; Bookheimer, 2002). Activation peaks lie for the most part in BA 44 or on its border with PMv, in proximity to several reported speech production peaks. Zatorre et al. (1996) suggested that left IFG activation might reflect processes involved in phonetic segmentation, since it was observed in studies where judgments required parsing a syllable into its constituent phonemes, but not in other studies where this was not required (e.g. Sergent et al., 1992). This hypothesis was tested and confirmed by Burton et al. (2000), who found left IFG activation only when explicit segmentation was required.

Several researchers have interpreted these activations as reflecting recruitment of articulatory codes in order to perform these tasks (e.g. Zatorre et al., 1992). However there are other possibilities: one is that left IFG activation might reflect cognitive aspects of the tasks which are not linguistic per se. Although Zatorre et al. (1996) reported left

Table 1.2 Phonological processing studies

#	Study	Method	*N*	Task	Control	Frontal peaks		
						x	*y*	*z*
1	Démonet et al. (1992)	PET*	9	Phoneme monitoring	Tone frequency analysis	–51	18	23
2	Zatorre et al. (1992)	PET*	10	Match final consonants	Passive words and pseudowords	–48	2	26
3	Démonet et al. (1994)	PET*	9	Phoneme monitoring	Tone frequency analysis	–40	3	31
4	Fiez et al. (1995)	PET*	10	Auditory target detection (speech and tones)	Fixation	–40	16	10
						41	22	6
5	Zatorre et al. (1996)	PET*	11	Phoneme monitoring	Passive words	–44	7	30
			11	Match final consonants	Passive words	–35	20	24
			10	Match final consonants	Noise	–43	4	30
			10	Match final consonants	Pitch discrimination	–57	5	32
6	Burton et al. (2000)	fMRI*	8	Segment and match initial consonants	Tone discrimination	–47	14	32
7	Poldrack et al. (2001)	fMRI	8	Rhyme judgment	Letter case judgement	–48	34	12
						–52	16	0
8	Burton & Small (2006)	fMRI*	10	Match initial consonants, match initial tones	Rest	–63	8	23

IFG activation for phoneme discrimination versus pitch discrimination, suggesting linguistic specificity, not all studies have reached such a conclusion. Fiez et al. (1995) did not find a difference between speech and tone processing in any frontal regions. Gelfand & Bookheimer (2002) found that the left posterior IFG was equally active for sequence-manipulation tasks performed on linguistic or non-linguistic stimuli, raising the possibility that this region has a general role in sequencing. However the linguistic units manipulated in that study were syllables rather than phonemes, giving the task more of a syntactic character, and possibly tapping an orthogonal function of the left IFG. Burton & Small (2006) found no difference in the left IFG between phonological and tone tasks which both involved segmentation and comparison of segmented units. But accuracy was significantly poorer on the tone task, so the presumably greater cognitive demands in the tone condition may have contributed to left IFG activity. Binder et al. (2004) employed a consonant discrimination task and used dissociations between accuracy and reaction time to identify regions that were involved in perceptual processes (argued to correlate with accuracy) and decision processes (argued to correlate with reaction time). Correlations with accuracy were found in superior temporal regions, supporting a role for these areas in sensory processing, whereas correlations with RT were reported bilaterally in the IFG. This would suggest that the IFG is involved with the decision process rather than phonetic discrimination, however the region activated was on the border of the operculum and the anterior insula, far ventral and clearly distinct from the IFG regions activated in other studies of phonological processing (see Figure 1.1).

The region termed the inferior frontal junction (IFJ), which lies at the junction of the inferior frontal sulcus and the inferior precentral sulcus, has recently been argued to be important for cognitive control, i.e. coordinating thoughts and actions according to internal goals (Brass et al., 2005). The location of the IFJ proposed by Brass et al. (2005) is shown in Figure 1.1: it is in the vicinity of activations in phonological processing studies, raising the possibility that these activations may reflect in whole or in part high level cognitive processes which are not specifically linguistic. This highlights the need for further phonological processing and phonetic discrimination studies which employ precisely matched nonlinguistic control tasks, where performance is matched and not at ceiling on either task.

Another possible explanation for left IFG activations which has been proposed is that they might reflect acoustic properties of the stimuli, such as rapid temporal transitions. However data on this issue is quite inconsistent. Fiez et al. (1995) were the first to propose that the left IFG might preferentially process stimuli with rapid temporal transitions; they reported that a frontal opercular region was more active for speech and nonspeech stimuli with rapid transitions, relative to vowels (which are steady state). Ventral frontal regions which preferred rapid over steady state or slower transitions have also been reported by Johnsrude et al. (1997) and Joanisse & Gati (2003), however the exact areas identified vary markedly from study to study. Moreover, Joanisse & Gati (2003) report, but do not discuss, a much more extensive left IFG region which responded more to steady state than rapidly changing stimuli, and the coordinates of this region (–49, 3, 16) place it much closer to the area typically identified in phonological processing studies than any of the more ventral regions which have been identified as preferring rapid transitions.

It is noteworthy that classic studies of phonological processing did not employ phonetically challenging tasks. Two recent studies have been exceptional in this respect. Blumstein et al. (2005) presented subjects with stimuli on a [ta]–[da] continuum where VOT varied between 0 ms and 40 ms. When boundary stimuli (20 ms VOT), which are ambiguous, were compared to endpoint stimuli (0 ms or 40 ms VOT), which are good exemplars of their respective categories, there was an extensive area of activation in the left IFG (–50, 26, 24; see Figure 1.1) This may reflect increased phonetic processing in an attempt to categorize the ambiguous stimulus, although it must be noted that RTs were longer for boundary stimuli, as expected for such tasks, so increased left IFG activity could also reflect greater duration of other cognitive processes. Liebenthal et al. (2005) used stimuli on a [ba]–[da] continuum as well as matched nonphonetic stimuli where formants were manipulated to create phonetically impossible sounds. Although discrimination of both kinds of stimuli activated inferior frontal regions, greater activity for speech over nonspeech was observed only in superior temporal areas. It is possible that the close acoustic match between the speech and nonspeech led the nonspeech stimuli to engage IFG regions involved in phonetic analysis, even though such analysis would ultimately not lead to a phonetic percept.

In sum neuroimaging studies have provided substantial evidence that the left IFG is involved in phonological processing, but have not conclusively ruled out that this might be due to task factors which are not strictly phonological.

Studies with aphasic patients, cortical excisions, electrical stimulation, and transcranial magnetic stimulation (TMS) have all revealed strong evidence for the involvement of motor and premotor regions in phonetic discrimination and identification tasks. We discuss now experiments using each of these methodologies.

Early aphasiologists held that speech perception was dependent on left superior temporal areas in the vicinity of primary auditory cortex (Wernicke, 1874; Luria, 1970). However the systematic investigation of phonetic perception in aphasic patients completely overturned this notion. Blumstein et al. (1977a) tested 25 patients who were roughly evenly split between Broca's, "mixed anteriors" (patients with anterior lesions but comprehension deficits), Wernicke's, and other patients with posterior lesions. The main task used was same/different discrimination on pairs of words or pseudowords differing in single phonemes. The mixed anterior patients performed the worst, followed by Wernicke's and other posterior patients, with Broca's patients performing the best. This was a surprising result, because the Wernicke's aphasics had the worst comprehension scores, yet they outperformed the mixed anterior patients who had only moderate comprehension deficits. The authors argued that comprehension deficits in Wernicke's aphasia could not be explained solely as a consequence of a phonetic perceptual deficit. After exclusion of the Broca's aphasics who performed well on phonetic discrimination as well as having good comprehension, there was no significant correlation between phonetic discrimination and comprehension among the remaining patients. This demonstrates that explicit phonetic perception tasks do not necessarily tap into the same neural resources responsible for speech perception in more naturalistic settings (Hickok & Poeppel, 2000, 2004). In another study, Blumstein et al. (1977b) tested 16 aphasic patients, plus control subjects, on phoneme discrimination and identification tasks using a [ta]–[da] VOT continuum. Most patients performed reasonably well on the discrimination task, but almost half failed on the identification task, even several who had been able to discriminate the phonemes. This study also confirmed the lack of relationship between the phonetic tasks and auditory language comprehension assessed more generally, and furthermore reported no link between impairments in production and perception of VOT.

Basso et al. (1977) studied a much larger group of 50 aphasic patients, asking them to classify stimuli on a [ta]–[da] VOT continuum. The lack of association between performance on a phonetic task and comprehension was confirmed: numerous patients performed poorly on the task despite having good comprehension, and some patients with poor comprehension succeeded on the task. However Basso et al. (1977) found strong evidence for links between speech perception and production, unlike Blumstein et al. (1977b). Out of the 21 nonfluent patients, 20 were impaired on this task (11 severely). In contrast, only 17 of 29 fluent patients were impaired (8 severely). Furthermore, when patients were classified according to whether or not they made phonemic errors in speech production (phonemic paraphasias, conduites d'approche, neologisms), a reliable association was found between these output deficits, and poor performance on the VOT classification task. A later study of 69 patients also supported a link between phonemic output disorders and the ability to discriminate between meaningless consonant-consonant-vowel-consonant (CCVC) syllables (Miceli et al., 1980).

None of these studies reported any information regarding lesion localization besides a gross distinction between anterior and posterior sites. A study by Gainotti et al. (1982) provided plots of lesion location for 60 aphasic patients who were tested on the same CCVC discrimination task of Miceli et al. (1980), as well as three further comprehension measures. When considering only patients with lesions to a single lobe, frontal patients performed by far the worst on the phoneme discrimination test (mean 9.33), followed by parietal patients (mean 4.86) and temporal patients (mean 2.00). A later study found that patients with deficits in phonetic processing tended to have left parietal lesions (Caplan et al., 1995), but the number of subjects (10) was much less than in the study by Gainotti et al. (1982).

Two conclusions can be drawn from studies of phonetic discrimination and identification in aphasic patients. First, phonetic deficits tend to be associated with frontal lesions, with nonfluent aphasia, and with disorders involving phonemic output. This is highly suggestive of motor system involvement in these tasks. But secondly, there is a relatively poor relationship between phonetic tasks and comprehension assessed by other means. This implies that although motor and premotor areas may be crucial for performing phonetic tasks, they might be dispensable for comprehension in more natural situations (Hickok & Poeppel, 2000, 2004).

One limitation of studies of aphasic patients is that lesions are typically large, and their boundaries and possible effects via diaschisis are difficult to define. In contrast, cortical excisions to remove epileptogenic tissue or tumors produce circumscribed, clearly defined lesions. However there are very few reports of speech perception deficits subsequent to cortical excisions. Taylor (1979) described deficits in phoneme discrimination in a series of patients with epileptogenic zones in the central (Rolandic) regions for the face. Patients were asked to discriminate 108 phonemes embedded in nonsense words. Left central patients performed very poorly even before surgery (mean 45.6 correct), and even worse after surgery (mean 30.6 correct). In follow-up testing they improved slightly (mean 37.4), which probably represents the chronic level of impairment. Right central patients performed much better, as did left and right frontal patients; these groups had scores in the 60s and low 70s before and after surgery. Temporal patients also had scores in the 70s. These remarkable findings provide evidence for a crucial role for the left motor/premotor face region in phonetic perception. The main outstanding question is to what extent these severe deficits would impact speech perception under normal conditions, since the studies with patients reviewed above showed that task performance and comprehension are not highly correlated. Left central patients were also impaired in word fluency (generating a set of words related to a prompt). This is not surprising given that this is the region that most reliably produces speech arrest, as reviewed above.

Another patient had surgical excision to remove a tumor in Broca's area, and was tested on final consonant and tone matching tasks (Zatorre et al., 1992). The patient performed at chance on the phonological task, but above chance on the tone task. These observations are consistent with the findings of Taylor (1979), and also provide evidence that the deficit may be linguistically specific to some degree. Finally, auditory comprehension deficits have been observed following resection of tumors involving premotor cortex (H. Duffau, personal communication, July 18, 2006). These deficits have not been systematically studied to date as the group has focused on the role of PMv and PMd in speech production (Duffau et al., 2003).

Direct electrical stimulation of the cortex has also produced transient disturbances in the ability to discriminate speech sounds (Ojemann & Mateer, 1979; Ojemann, 1981, 1983). Sites affecting speech perception have been found in inferior frontal regions, superior temporal cortex, and the inferior parietal lobe. There was a remarkable association between sites involved in discrimination of stop consonants and those involved in mimicry of sequences of orofacial movements. 86% of orofacial sites also produced phonetic deficits, and 81% of phonetic sites produced orofacial deficits. Stimulation of some sites led to deficits even in the ability to imitate a single orofacial movement; these sites were found exclusively in posterior inferior frontal cortex and are presumably motor or premotor sites. One other electrical stimulation study has reported phoneme identification errors following stimulation of inferior frontal sites (Boatman, 1995), though several others have not (for review see Boatman, 2004).

Two TMS studies have demonstrated effects on phonological or phonetic tasks as a result of stimulation to frontal sites. Gough et al. (2005) found that subjects were significantly slower to make homophone judgments (e.g. do *ate* and *eight* sound alike?) when stimulated with three pulses in the posterior left IFG. Stimulation to a more anterior site did not affect this phonological task. A semantic task showed the opposite pattern (slowing when the anterior site was stimulated), consistent with the parcellation of the left IFG into posterior phonological and anterior semantic regions which has been proposed (Poldrack et al., 1999; Bookheimer, 2002).

In a study which tested phonetic perception directly, Meister et al. (submitted) used offline rTMS to create “virtual lesions” in three different locations: left premotor cortex, left superior temporal gyrus, or the vertex (a control site). There was also a baseline condition with no stimulation. Participants performed two tasks: stop discrimination in noise, and subtle color discrimination. The difficulty of both tasks was equated for each subject independently using an adaptive staircase procedure prior to the main part of the study. The results showed that rTMS to premotor cortex had a dramatic impact on performance of the phonetic task, with increases in the error rate in all 8 subjects. Stimulation of the STG had a smaller effect. The color task was unaffected by stimulation to any site. This study demonstrates that left premotor cortex is crucial for a pure phonetic perception task. The premotor site stimulated was a relatively superior region activated in previous studies of speech perception (Wilson et al., 2004), which is distinct from the more ventral premotor regions stimulated in electrical stimulation studies (e.g. Ojemann, 1983), although it may be encompassed in the areas excised in some patients described by Taylor (1979).

In sum there is a great deal of evidence for the involvement of left frontal regions in phonological processing and phonetic perception tasks. Lesion and stimulation studies complement functional imaging studies by showing that these areas are not merely activated; they are in fact necessary for task performance. There are two main limitations to this body of work. The first is that nonlinguistic control conditions have not always been employed, or have been insufficient. However some studies with well matched control conditions have reported differential involvement of left frontal regions in linguistic tasks (Zatorre et al., 1992, 1996; Meister et al., submitted). The second major limitation is that studies of aphasic patients have conclusively demonstrated that there is little to no relationship between performance on phonetic discrimination or identification tasks, and more general measures of comprehension. While poor comprehension despite intact speech perception would reflect higher level linguistic deficits, good comprehension in the face of impaired phonetic perception is somewhat paradoxical. Three explanations can be identified (Blumstein et al., 1977b). First, phonetic tasks often restrict the range of phonetic cues available (e.g. when synthetic phonemes are used); it might be the case that patients do much better when multiple cues are available. Secondly, metalinguistic awareness of segments is typically necessary to perform phonetic tasks but not for normal comprehension; these metalinguistic processes may be selectively impaired in some patients. Thirdly, normal language comprehension provides rich syntactic, semantic and discourse contextual top-down information. Patients may be able to use these nonphonetic cues to substitute for poor phonetic perception. A useful paradigm for examining this possibility has recently been proposed (Fink et al., 2006). Patients were tested for phonetic perception where there was one key word embedded in sentences which either (i) contained morphological or syntactic cues; (ii) were meaningful but did not contain such cues; (iii) were meaningless. Aphasic patients performed much better on the first class of sentence, whereas they were equally impaired on meaningful and meaningless sentences which did not contain any informative cues. The fact that meaningful sentences without cues did not offer any benefit suggests that dissociations between phonetic tasks and comprehension may not be a consequence of an ecologically valid context per se, but rather result from the post-phonetic cues provided by such a context. This supports the view that the role for motor and premotor cortex demonstrated in studies of phonetic tasks probably generalizes to a role in speech perception more generally.

1.4 Effortful extraction of phonetic cues

A number of functional neuroimaging studies have investigated speech perception under non-optimal conditions such as temporal compression (Poldrack et al., 2001), noise (Callan et al., 2003; Binder et al., 2004; Scott et al., 2004, Zekveld et al., 2006) or a range of degraded conditions (Davis & Johnsrude, 2003). These studies are important because they permit the testing of the hypothesis that motor regions are increasingly important in speech perception under conditions when the perceptual input is degraded (Callan et al., 2004).

Frontal activation peaks from five of these studies are plotted in Figure 1.1, and the studies are characterized in Table 1.3. Poldrack et al. (2001) reported a contrast identifying "convex" responses to temporally compressed speech whereby compressed but intelligible speech produced more activation than either unaltered speech or speech that was so temporally compressed that it was unintelligible. Activation peaks were reported on the border of BA 44 and BA 45, and more medially in the frontal operculum and anterior insula. Similarly Davis & Johnsrude (2003) presented sentences at various levels of distortion, and identified regions which showed increased responses to distortion once intelligibility was factored out. Peaks were reported in the ventral IFG as well as anterior to BA 6, but the area activated spanned much of dorsal and ventral premotor cortex and the posterior IFG. Peaks in the ventral IFG were reported by Binder et al. (2004) and Scott et al. (2004) for processing speech in noise. The same colored symbol has been used in Figure 1.1 to denote the study of Blumstein et al. (2005) discussed

Table 1.3 Studies of speech comprehension under phonetically challenging conditions

#	Study	Method	*N*	Contrast	Frontal peaks		
					x	*y*	*z*
1	Poldrack et al. (2001)	fMRI	8	Convex response to compression	–52	22	12
					–34	14	20
					36	26	8
					–38	34	8
2	Davis & Johnsrude (2003)	fMRI	12	Increase for distorted speech relative to clear speech and SCN	–48	14	–6
					–40	2	44
3	Binder et al. (2004)	fMRI*	18	Negative correlation with accuracy	–45	10	1
					43	10	6
					–48	40	12
4	Giraud et al. (2004)	fMRI	8	Auditory search for phonetic cues	–56	10	30
5	Scott et al. (2004)	PET	7	Speech-in-noise versus speech-in-speech	–34	28	44
				Negative correlation with SNR	–42	20	–12
					–14	2	70
6	Blumstein et al. (2005)	fMRI*	12	Stimuli on VOT categorical border	–51	26	27

above, since the stimulus on the phonetic boundary may also tax phonetic identification processes in a similar way.

Also indirectly related is an elegant study by Giraud et al. (2004) in which subjects were presented with complex speech envelope noises which were initially not understood, but which could be understood after an intervening training session. The contrast of particular interest here identified regions involved in “auditory search” as those which were more active once subjects were aware that some of the stimuli contained linguistic information. The region activated by this contrast was located in dorsal BA 44, immediately anterior to peaks for orolaryngeal motor control.

Taken together these studies suggest that speech perception in noise does lead to additional recruitment of frontal areas, but the exact areas implicated vary quite considerably from study to study.

If the motor system makes a greater contribution to perception under non-optimal conditions, then we might expect non-fluent aphasic patients to show poorer comprehension in noise. In support of this, Moineau et al. (2005) found that whereas Broca's aphasics performed quite well at a lexical comprehension task under normal conditions, they were differentially impaired relative to control subjects when the words were subjected to acoustic degradation (a combination of low pass filtering and temporal compression). An important avenue for future research will be to determine whether comprehension performance under acoustically challenging conditions shows a better correlation with performance on phonetic tasks than does regular comprehension.

1.5 Passive speech perception

A large number of functional imaging studies have described neural activity when subjects perceive speech at various linguistic levels—syllables, words, sentences or narratives—with few or no task demands (for a previous review, see Indefrey & Cutler, 2004). This body of work provides an important perspective which complements the findings discussed above based on explicit phonological or phonetic tasks. However there are at least two major limitations to studies with few or no task demands. Firstly, the nature of the processing performed by subjects is unconstrained and often unknown. For instance, if subjects are presented with single words in isolation, it can be difficult to determine whether resultant neural activity reflects phonetic perception, lexical access, semantic processing, or so on. Individuals may also differ in the kinds of processing they perform. A second major limitation relates to the necessity of comparing speech-related activity to some control condition. Regardless of whether the control condition is a resting state, or whether control stimuli are used which match various acoustic properties of speech, it is clear that cognitive processes will take place during the control condition which are difficult to characterize (Shulman et al., 1997; Binder et al., 1999; Gusnard and Raichle, 2001; McKiernan et al., 2003).

Activation peaks from studies at four different levels—syllables (and pseudowords), words, sentences and narratives—are plotted in Figures 1.2 and 1.3. This is intended to be a reasonably exhaustive summary of such studies, although certainly there must be other published studies which have reported relevant contrasts. Note that for these studies, color is used to denote the linguistic level, and the shape of the markers specifies the kind of control condition. We now discuss the pattern of findings at each of these four levels in turn.

Studies where participants listened to syllables or pseudowords are listed in Table 1.4 and plotted with white markers in Figure 1.2. The majority of studies have reported activations in lateral frontal regions, and it can be noted that many of the studies which did not had rather small numbers of subjects (e.g. Wise et al., 1991; Fiez et al., 1996). Activation peaks were quite consistently observed in BA 6 and its vicinity. These activations were bilateral in some studies (Binder et al., 2000; Wilson et al., 2004; Uppenkamp et al., 2006; Wilson & Iacoboni, 2006), and left-lateralized in others

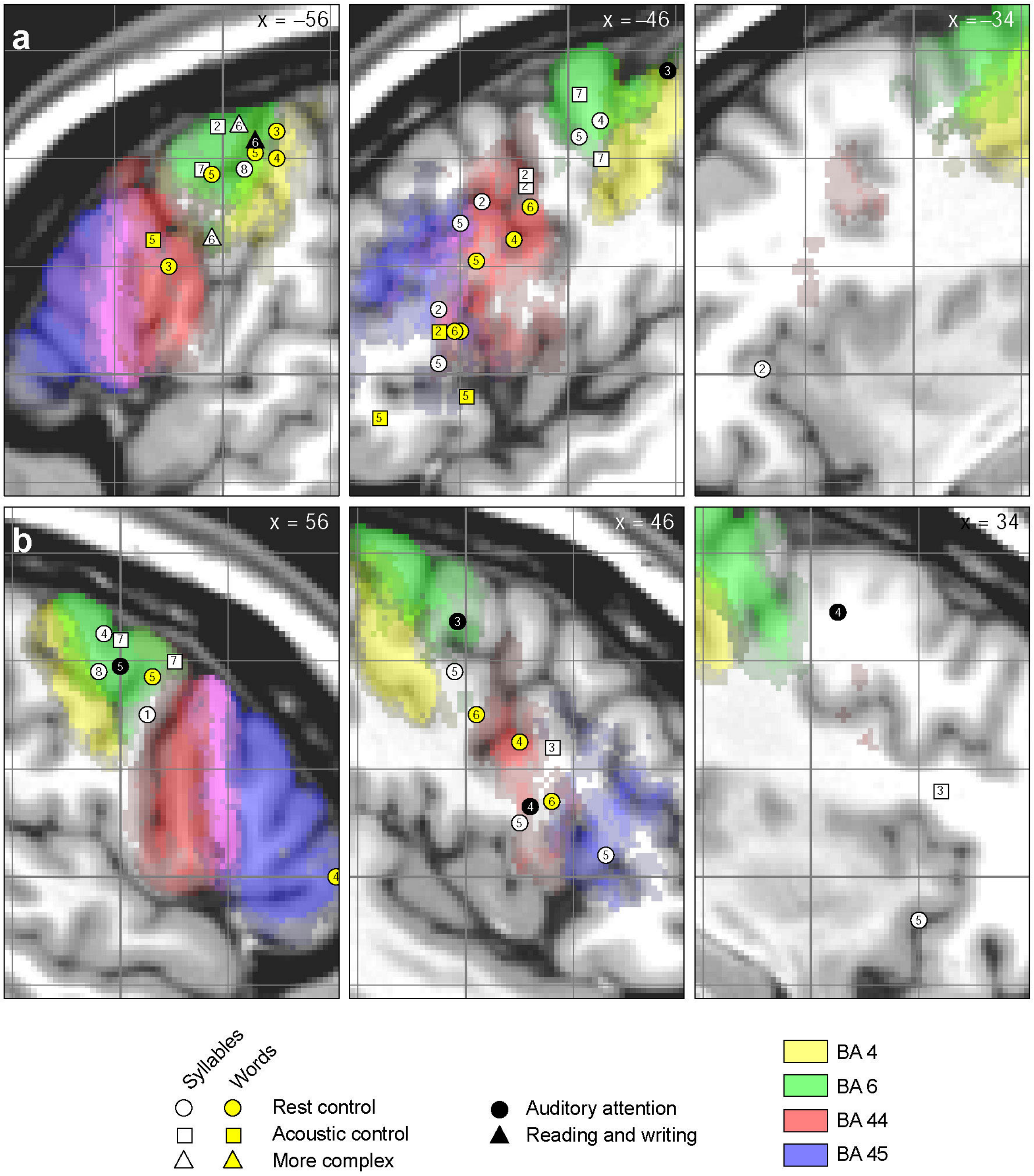

Figure 1.2 Activation peaks from studies of syllable and pseudoword perception, and comprehension of single words in isolation, in left (a) and right (b) lateral frontal regions and the insula. Also shown are studies of auditory attention (Zatorre et al., 1999: black circle 3; Lipschutz et al., 2002: black circle 4; Degerman et al., 2006: black circle 5), and reading and writing single letters (Longcamp et al., 2003: black triangle 6). The syllable and word perception studies are listed in Tables 1.4 and 1.5. See caption to Figure 1.1 for additional information.

Table 1.4 Syllable and pseudoword perception studies

#	Study	Method	*N*	Stimuli	Task	Control	Frontal peaks		
							x	*y*	*z*
1	Binder et al. (2000)[a]	fMRI*	28	Pseudowords	None	Rest	–52	–12	43
							43	9	28
							52	4	33
				Pseudowords	None	Tones			None
2	Benson et al. (2001)	fMRI$_{ss}$	12	Syllables	None	Rest	–47	24	12
							–43	16	32
							–40	28	1
				Syllables	None	Chords	–47	8	35
							–49	8	37
							–54	1	46
3	Vouloumanos et al. (2001)	fMRI	15	Pseudowords	Identify from train of tone standards	Same for complex nonspeech	40	24	16
				Pseudowords	Identify from train of tone standards	Same for complex nonspeech	48	16	24
4	Wilson et al. (2004)	fMRI	10	Syllables	None	Rest	–50	–6	47
							55	–3	45
5	McNealy et al. (2005)	fMRI	27	Syllables	None	Rest	–46	24	2
							–48	20	28
							–50	–2	44
							44	10	10
							48	26	4
							38	20	–8
							48	–2	38
6	Pulvermüller et al. (2006)	fMRI$_{ss}$	12	Syllables [pV]	None	ROI analysis	–54	–3	46
				Syllables [tV]	None	ROI analysis	–60	2	25
7	Uppenkamp et al. (2006)	fMRI$_{ss}$	10	Vowels (natural and synthetic)	Detect intensity change	Same for matched nonspeech	–50	–2	52
							–48	–6	40
							–56	4	38
							58	0	44
							58	10	40
8	Wilson & Iacoboni (2006)	fMRI	12	VCV with nonnative consonants	None	Rest (event-related)	–62	–4	38
							56	–4	38

Studies with no lateral frontal activation peaks					
Wise et al. (1991)	PET	6	Pseudowords	None	Rest
Fiez et al. (1996)	PET	3	Pseudowords	None	Rest
Bookheimer et al. (1998)	PET	8	Pseudowords	None	Rest
Giraud et al. (2000)	PET	6	Vowels	None	Expecting vowels
Jancke et al. (2002)[b]	fMRI$_{ss}$	21	Syllables	Detect target item	Rest
Suzuki et al. (2002)[b]	fMRI	10	Syllables	None	Rest
Rimol et al. (2005)	fMRI$_{ss}$	17	Consonants	Detect repeated items	Same for matched noise
			Syllables	Detect repeated items	Same for matched noise
Rimol et al. (2006)	fMRI	18	Syllables	Detect repeated items	Same for tones

Note. In this and subsequent tables, the subscript *ss* denotes that sparse sampling was used and stimuli were presented in silent periods between scans.

[a] Coordinates for frontal activations were not reported and were estimated based on the figures provided.

[b] The field of view probably excluded many frontal regions.

(Benson et al., 2001; McNealy et al., 2006). Most of these studies used resting control conditions for the relevant contrasts except for Benson et al. (2001) where the contrast was against chords, and Uppenkamp et al. (2006) where an acoustically matched nonspeech stimulus was used. Wilson et al. (2004) showed in a region-of-interest (ROI) analysis that the region in question responded more to speech than to noise or an environmental sound. One study which found activations in the vicinity of BA 6 for pseudowords versus rest did not report any activity when pseudowords were compared to tones (Binder et al., 2000). Lateral frontal activations outside of this region were sporadic and inconsistent except in two studies: Benson et al. (2001) where it is worth noting that about 80 mins of functional data were collected per subject, and McNealy et al. (2006) who presented concatenated syllables as an “alien language” which probably encouraged the recruitment of higher level linguistic processes such as attempts at word segmentation.

One recent study examined responses to passive perception of alveolar [t] and bilabial [p] consonants in motor regions of interest which were defined by the production of tongue and lip movements, or production of the same phonemes (Pulvermüller et al., 2006). For the perception condition, a significant interaction of place of articulation by ROI was found for a pair of ROIs which were defined anterior to areas activated by the tongue and lip movement tasks. The lip ROI had coordinates (–56, –8, 46) and the coordinates of the tongue ROI were (–60, –10, 25). Furthermore a gradient was observed in the precentral gyrus more generally whereby listening to or producing tongue-related phonemes activated more ventral regions and listening to or producing lip-related phonemes activated more dorsal regions, in line with the arrangement of these two effectors (Penfield & Roberts, 1959). The phoneme-specific responses observed in this study suggest that frontal activations to passive speech perception reflect articulatory codes rather than any more general perceptual or cognitive process. However the study is not entirely compelling empirically: although it is well designed, the definitions of the ROIs appear to be somewhat ad hoc.

Studies where participants listened to isolated words are listed in Table 1.5 and plotted with yellow markers in Figure 1.2. Approximately half of all studies reported one or more activation peaks in lateral frontal areas. Three studies reported activations in the same BA 6 region activated by listening to syllables, two on the left hemisphere (Binder et al., 1996; Binder et al., 2000) and one bilaterally (Specht & Reul, 2003). All of these were contrasts relative to rest. Activation peaks in left BA 44 were reported in four studies (Binder et al., 1996; Binder et al., 2000; Specht & Reul, 2003; Hickok & Okada, 2006), and a further three peaks were observed more ventrally on the border of BA 44 and BA 45 (Mellett et al., 1996; Price et al., 1996; Okada & Hickok, 2006). Right hemisphere activations were observed only occasionally and not in any consistent location.

For studies of sentence comprehension, it is unavoidable that active conditions will involve semantic comprehension of the sentence, whereas an acoustically matched control condition, if used, will not. Therefore we included in our review studies employing relatively easy tasks which tapped sentence comprehension, for instance intelligibility judgment (Davis & Johnsrude, 2003) or assessing relatedness of a probe word (Rodd et al., 2005). The majority of studies reported lateral frontal activations, and the locations of these activations were quite consistent (Table 1.6, Figure 1.3). The largest cluster of peaks was centered around the pars triangularis of the IFG, falling rather ventrally on the border of BA 44, BA 45, and presumably also BA 47, for which no probabilistic cytoarchitectonic map is available (Schlosser et al., 1998; Humphries et al., 2001; Davis & Johnsrude, 2003; Ben-Shacher et al., 2004; Meyer et al., 2004; Rodd et

Table 1.5 Isolated word comprehension studies

#	Study	Method	*N*	Task	Control	Frontal peaks		
						x	*y*	*z*
1	Mellet et al. (1996)	PET*	9	None	Rest	–46	20	10
2	Price et al. (1996)[a]	PET	4	None	Reversed words	–48	24	8
		PET	6	None	Rest	None		
		PET	6	None	Rest	None		
3	Binder et al. (1996)[b]	fMRI*	12	Press button on block onset	Rest, same task	–52	9	22
						–52	–13	48
				Press button on block onset/offset	Tones, same task	None		
4	Binder et al. (2000)[c]	fMRI*	28	None	Rest	–52	–12	43
						–43	9	28
						43	9	28
						52	41	2
				None	Tones at same rate	None		
				None	Pseudowords at same rate	None		
				None	Reversed words at same rate	None		
5	Specht & Reul (2003)	fMRI	12	None	Rest (event-related)	–44	17	21
						51	6	37
						–55	–6	41
						–55	2	37
				None	Environmental sounds	–44	19	–4
				None	Tones	–48	19	–4
						–44	35	–8
						–51	13	25
6	Okada & Hickok (2006)[d]	fMRI*	10	None	Rest (event related)	–44	21	10
						–46	6	34
						42	16	16
						49	1	33
	Studies with no lateral frontal activation peaks							
	Petersen et al. (1988)	PET	17	None	Rest			
	Petersen et	PET	17	None	Rest			

al. (1989)				
Fiez et al. (1996)	PET	3	None	Rest
Mummery et al. (1999)	PET	6	None	SCN at same rates
Giraud et al. (2000)	PET	6	None	Vowels
Ashtari et al. (2004)	fMRI	9	Judge pairs same/different	Tone pairs, same task
Rimol et al. (2006)	fMRI	18	Detect repeated items	Tones, same task

[a] Results are provided for three separate experiments, numbered 3, 4 and 5 in the original paper.

[b] Coordinates for frontal activations were not reported and were estimated based on the figures provided.

[c] Coordinates for frontal activations were not reported and were estimated based on the figures provided.

[d] Auditory words were presented along with pictures.

al., 2005; Dehaene-Lambertz et al., 2006). Two further studies reported activations more medially, in the anterior insula (Meyer et al., 2004; Zekveld et al., 2006). A second cluster of activations were found more dorsally in BA 44 (Humphries et al., 2001; Ben-Shacher et al., 2004; Meyer et al., 2004; Humphries et al., 2006; Zekveld et al., 2006). No studies reported any activations more dorsally, such as in the BA 6 region activated by listening to syllables and words. Some studies reported right hemisphere peaks in BA 44 and BA 45, however they tended to be less consistently localized than activations in the left hemisphere (see Table 1.6 for references).

Studies in which subjects listened to narratives are listed in Table 1.7, and activation peaks are shown in olive in Figure 1.3. Setting aside for a moment the study of Wilson et al. (submitted; chapter 4), peaks have been reported in a very wide range of left inferior

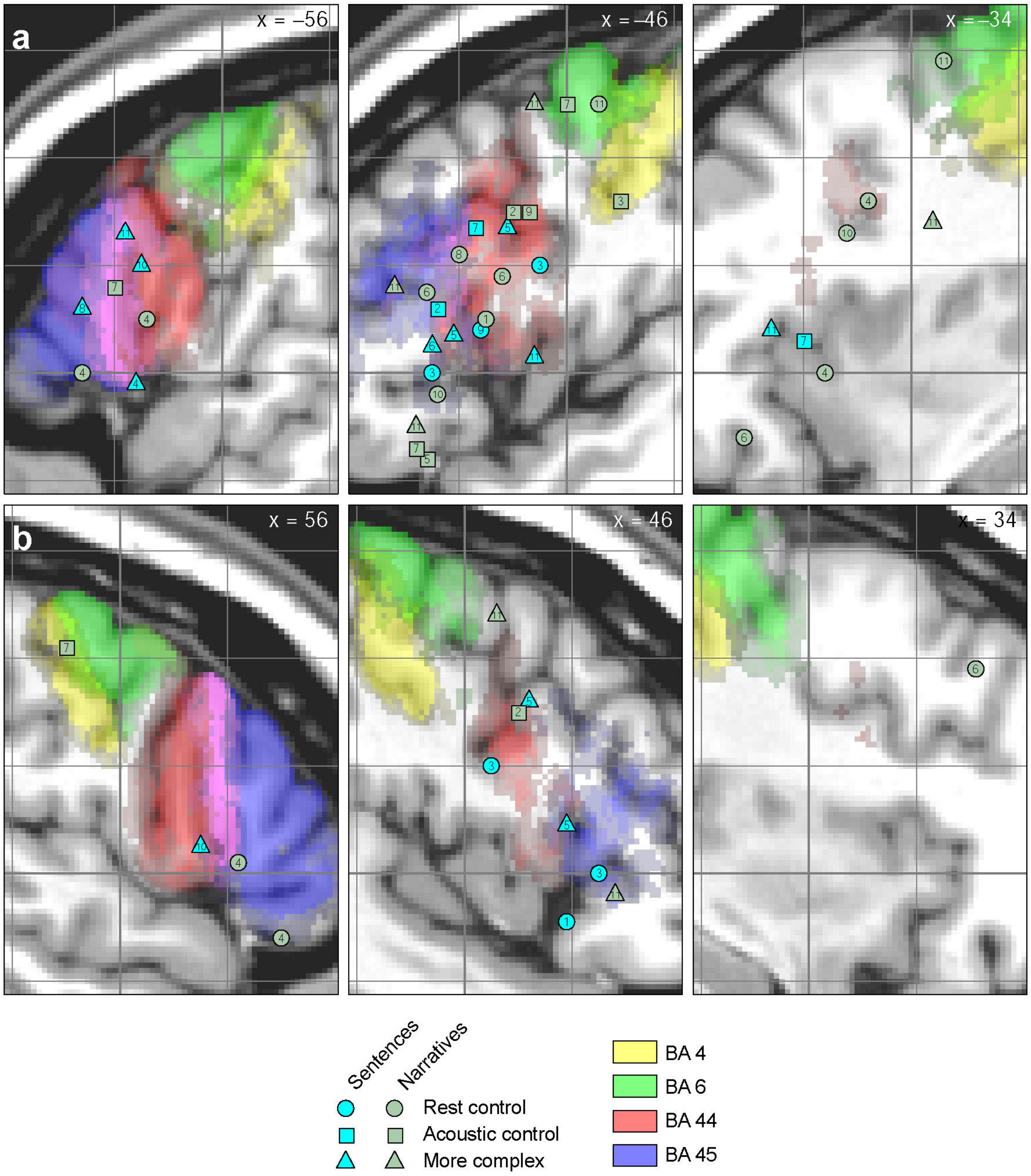

Figure 1.3 Activation peaks from studies of sentence and narrative comprehension, in left (a) and right (b) lateral frontal regions and the insula. Studies included are listed in Tables 1.6 and 1.7. See caption to figure 1.1 for additional information.

Table 1.6 Sentence comprehension studies

#	Study	Method	*N*	Task	Control	Frontal peaks		
						x	*y*	*z*
1	Müller et al. (1997)	PET*	9	None	Rest	48	20	–10
2	Schlosser et al. (1998)	fMRI*	14	None	Sentences in unknown language	–47	24	14
3	Humphries et al. (2001)[a]	fMRI*	7	None	Rest	–51	26	1
						–51	4	22
						51	26	1
						51	4	22
				None	Meaningful sequences of environmental sounds	None		
4	Davis & Johnsrude (2003)[b]	fMRI$_{ss}$	12	Rate intelligibility of distorted sentences	N/A—correlation with intelligibility	–58	16	–2
5	Ben-Shacher et al. (2004)	fMRI*	11	Probe question (excluded from data analysis)	Rest (quasi-event-related)	–43	21	9
						48	19	11
						–41	10	30
						44	11	35
6	Meyer et al. (2004)	fMRI*	14	None in analyzed data	Unintelligible degraded sentences	–47	25	7
7	Meyer et al. (2005)	fMRI*	12	Detect nonsense syllables	Laugher, same task	–43	16	30
						–40	20	8
8	Rodd et al. (2005)	fMRI$_{ss}$	15	Judge probe word relatedness	Noise, button press according to word prompt	–54	26	12
		fMRI$_{ss}$	15	None	Noise	None		
9	Dehaene-Lambertz et al. (2006)	fMRI	10	Listen	Rest (event-related)	–48	16	8
10	Humphries et al. (2006)[c]	fMRI	21	Rate meaningfulness	Rest (event-related)	–52	15	20
						52	15	5
				Rate meaningfulness	Same task on word lists	None		
11	Zekveld et al.	fMRI$_{ss}$*	10	Sentences in noise,	Noise, guess at task	–51	6	4

(2006)[d]			choose matching written sentence		–40	26	10
					–52	17	29
Studies with no lateral frontal activation peaks							
Wong et al. (1999)	PET	5	Press button on alternate stimuli	Reversed sentences, same task			
Giraud et al. (2000)	PET	6	None	Vowels			
Scott et al. (2000)	PET	8	None	Spectrally rotated sentences			
Wong et al. (2002)	PET	5	Detect repeated items	Reversed sentences, same task			
Narain et al. (2003)	fMRIss	11	None	Spectrally rotated sentences			
Peelle et al. (2004)	fMRIss	8	Semantic judgement	Unintelligable filtered sentences, judge speaker gender			
Humphries et al. (2005)[e]	fMRI	12	None	Rest			

[a] Coordinates for frontal activations were not reported and were estimated based on the figures provided.

[b] Sentences were degraded to different degrees with various kinds of distortion.

[c] Only inferior frontal regions seem to have been included in the field of view.

[d] Sentences were presented in noise at a range of signal to noise ratios.

[e] The field of view probably excluded many frontal regions.

frontal regions including BA 44 (Mazoyer et al., 1993; Dehaene et al., 1997; Papathanassiou et al., 2000; Tzourio-Mazoyer et al., 2004; Schmithorst et al., 2005), more medially in the inferior frontal junction (Papathanassiou et al., 2000; Alho et al., 2006), the border between BA 44 and BA 45 (Tzourio-Mazoyer et al., 2004; Crinion & Price, 2005; Skipper et al., 2005) and the IFG pars orbitalis (likely BA 47) (Crinion et al., 2003; Tzourio-Mazoyer et al., 2004; Crinion & Price, 2005; Alho et al., 2006). Two studies reported more dorsal peaks as well: Perani et al. (1998) found activation in left primary motor cortex, and Crinion & Price (2005) reported a peak in left BA 6 in the

Table 1.7 Narrative comprehension studies

#	Study	Method	N	Task	Control	Frontal peaks		
						x	y	z
1	Mazoyer et al. (1993)[a]	PET	16	None	Rest	–50	15	10
2	Dehaene et al. (1997)[b]	fMRI	8	None	Backwards unknown language	–50	10	30
						50	10	30
3	Perani et al. (1998)[c]	PET*	21	None	Backwards unknown language	–51	–12	34
4	Papathanassiou et al. (2000)[d]	PET	8	None	Rest	–56	26	0
						–60	14	10
						–36	16	0
						–40	8	32
						56	30	–12
						62	22	2
5	Crinion et al. (2003)	PET	17	None	Backwards speech	–44	26	–16
6	Tzaorio-Mazoyer et al. (2004)[e]	PET	20	None	Rest	36	31	38
						–37	51	–8
						37	51	–8
						–48	12	18
						–47	26	15
						–39	31	–12
7	Crinion & Price (2005)	fMRI	18	None	Backwards speech	–56	20	16
						–42	28	–14
						–46	0	50
						56	–10	42
8	Skipper et al. (2005)[f]	fMRI*	9	None	Rest	–51	19	25
9	Schmithorst et al. (2006)[g]	fMRI*	313	None	Tones	–42	6	33
10	Alho et al. (2006)	PET	10	None	Rest	–38	12	26
						–48	24	–4
11	Wilson et al. (submitted)	fMRI	12	None	Rest	–38	–6	58
						–46	–6	50
					N/A—intersubject correlation	–50	28	–10
						48	28	–4
						–46	32	16

					40	46	10
					–40	–4	28
					–44	6	50
					46	6	48
Studies with no lateral frontal activation peaks							
Giraud et al. (2000)	PET	6	None	Vowels			
Ahmad et al. (2003)[h]	fMRI	15	None	Backwards speech			

[a] Coordinates were not reported and were estimated based on the figures provided.

[b] Only results for listening in L1 are considered here. Coordinates were not reported and were estimated based on the figures provided.

[c] The frontal activation was found in individuals listening to a second language which they had acquired early and had high proficency with.

[d] Conjunction analysis.

[e] Coordinates were not reported but ROIs used have been published so their approximate centers were used here.

[f] A much more extensive set of left and right frontal areas were activated for listening to audiovisual speech versus rest.

[g] Subjects were children. The activation peak is from an Independent components analysis. The threshold in the random effects analysis was set extremely high and no frontal activity was observed.

[h] Subjects were children.

vicinity of activations there for syllable and single word studies discussed above. Few studies have reported activation in right frontal areas, and locations have varied considerably across studies (see Table 1.7 for references).

Wilson et al. (submitted; chapter 4) used intersubject correlational analysis (Hasson et al., 2004) to identify brain regions that were correlated across subjects as the subjects listened to the same set of narratives. An extensive network of bilateral inferior frontal and premotor regions were correlated across subjects, demonstrating that these regions are sensitive to time-varying properties of the linguistic input, or the processing that it invokes. Peaks were reported in the left and right IFG pars orbitalis, left and right dorsal premotor cortex, left inferior frontal junction (Figure 1.3) and more anteriorly in the right inferior frontal sulcus (not shown). However these were local maxima in extensive bilateral activations; essentially all of the regions containing activation peaks in previous studies showed significant intersubject correlations.

Several TMS studies have also investigated motor activity during passive speech perception. Fadiga et al. (2002) stimulated tongue motor cortex and recorded motor evoked potentials (MEPs) from the anterior tongue muscles. Subjects were presented with words and pseudowords containing alveolar trills (produced with the tongue), bilabial fricatives (produced with the lips), and tones. MEPs recorded during listening to trills were significantly larger than those recorded in other conditions, suggesting greater levels of endogenous motor activity in tongue motor cortex evoked by sounds produced with the tongue. Although this study provides some evidence that the motor activity is phoneme-specific, a more convincing demonstration would involve stimulating lip motor cortex and recording from the lip to confirm that the opposite results would be obtained.

The nature of this effect was further specified in another TMS study with a wider range of conditions (Watkins et al., 2003). Subjects were presented with continuous prose, environmental sounds, visual speech-related lip movements and visual eye and brow movements, while face motor cortex was stimulated and MEPs were recorded from the lip. Only listening to prose and viewing speech-related lip movements enhanced MEPs, and the effect was found only in the left hemisphere; there were no significant changes when the right hemisphere was stimulated. In a follow-up study, Watkins & Paus

(2004) argued that these changes in cortical excitability were mediated by Broca's area, because in a PET study Broca's area was the brain region most strongly correlated with cortical excitability.

Although TMS lacks the spatial resolution of fMRI or PET, these studies are important because they demonstrate directly that the motor system is facilitated by speech perception, whereas imaging studies can only infer this from colocalization of perception and production activations, which has rarely been explicitly shown.

1.6 Functional roles of frontal regions in speech perception

All of the activation peaks plotted so far are combined together in Figure 1.4. At a gross level, there is a clear distinction in the patterns of activation peaks observed in production and perception studies. Two speech motor regions—the anterior insula and DLPFC—have rarely been reported in speech perception studies, so probably do not play a significant role in perception. In the region spanning primary motor cortex, BA 6, PMv and BA 44, there is a general principle that speech production peaks are relatively posterior, and speech perception peaks are relatively anterior. Three broad regions can be identified as responsive to speech perception in a substantial number of studies: (1) dorsal BA 44 and adjacent inferior ventral premotor cortex (iPMv); (2) BA 6, anterior and dorsal to primary motor cortex; (3) a ventral region on the border of BA 44 and BA 45. We will now discuss each of these regions in turn.

Dorsal BA 44 and adjacent iPMv in the left hemisphere is the region most likely to contain articulatory representations of speech sounds which may be accessed in speech

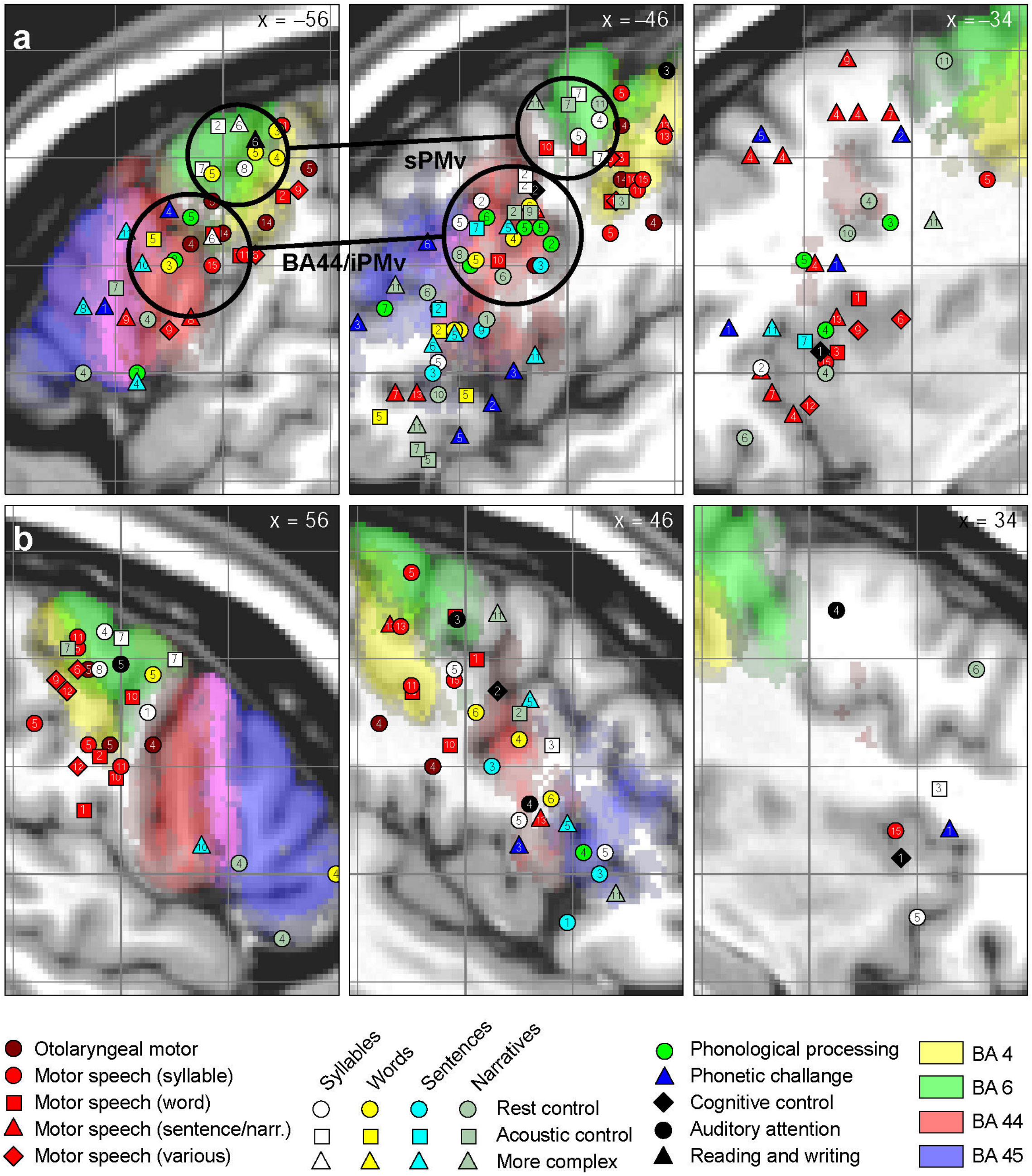

Figure 1.4 All activation peaks from Figures 1.1, 1.2 and 1.3. See captions to these figures for more information. The studies included here are listed in Tables 1.1 through 1.7, and in the captions to Figures 1.1 and 1.2. The approximate locations of two premotor regions argued to be important for speech perception are shown: dorsal BA 44/inferior ventral premotor cortex (iPMV), and superior ventral premotor cortex (sPMv).

perception along the lines suggested by the motor theory (Liberman et al., 1967; Liberman & Mattingly, 1985). Activations in this region were frequently reported in all levels of linguistic processing except for syllables (which tended to produce more dorsal activation peaks). Studies of phonological processing also yield activations which cluster quite tightly in this region (e.g. Zatorre, 1996), and three important studies of effortful phonetic processing produced activation in this vicinity (Poldrack et al., 2001; Giraud et al., 2004; Davis & Johnsrude, 2003). In support of an overlap with speech production, two production studies reported peaks which corresponded closely to many speech perception peaks (Turkeltaub et al., 2002; Kemeny et al., 2005). Moreover, the orolaryngeal motor control task employed by Braun et al. (1997), which was designed to involve all muscle groups used in speech in movements which were qualitatively similar to those used in speech, yielded two peaks in dorsal BA 44. A number of production studies with peaks in this vicinity observed them posterior to BA 44, in iPMv (Petersen et al., 1988; Fox et al., 2001; Turkeltaub et al., 2002; Wilson et al., 2004; Wilson & Iacoboni, 2006). It is possible that production peaks lie posterior to perception peaks because speech production involves actual projections to the speech musculature, and such neurons occur in greater concentrations more posteriorly in premotor cortex (He et al., 1993). A possible parallel is the case of motor imagery where activations for imagery have been reported to be anterior to those for movement (Ehrsson et al., 2003).

This region corresponds quite well to the area where deficits in phonetic discrimination as well as mimicry of orofacial movements were disrupted by direct electrical stimulation (Ojemann & Mateer, 1979; Ojemann, 1981, 1983). It is possible that the Rolandic lesions discussed by Taylor (1979) which lead to severe phonetic discrimination deficits may also have encroached on this area in many cases.

The second region frequently reported in speech perception studies was located in BA 6, anterior and dorsal to primary motor cortex for the mouth. This region was activated in many studies of syllable perception (Binder et al., 2000; Benson et al., 2001; Wilson et al., 2004; Uppenkamp et al., 2006; McNealy et al., 2006; Wilson & Iacoboni, 2006) and word perception (Binder et al., 1996; Binder et al., 2000; Specht & Reul, 2003), and two studies of narrative perception (Crinion & Price, 2005; Wilson et al., submitted).

We will refer to this region as superior ventral premotor cortex (sPMv). The strongest argument that it constitutes a sector of ventral premotor cortex despite its relatively superior location is that dorsal premotor cortex (in the macaque) does not contain any orofacial representations (Raos et al., 2003), whereas sPMv is strongly activated by speech production (Wilson et al., 2004). Recently it has been suggested that the dividing line between dorsal and ventral premotor cortex in humans is approximately $z = 51$ (Rizzolatti et al., 2002). The majority of sPMv activations are ventral to that plane, which provides further justification for the classification. Furthermore, lateral frontal activity in simple motor tasks presumed to reflect dorsal premotor cortex (PMd) is considerably medial and dorsal to sPMv (Fink et al., 1997; Kollias et al., 2001).

In the rTMS study of Meister et al. (submitted), stimulation of functionally-defined sPMv led to significant deficits in the ability to discriminate stop consonants in noise. This demonstrates that sPMv is crucial for speech perception, at least in the context of an overt task.

In previous studies we suggested that sPMv might be involved in articulatory representations of perceived speech (Wilson et al., 2004; Wilson & Iacoboni, 2006). This is possible, although most sPMv peaks are dorsal and anterior to peaks for either speech production or movement of the articulators. It is also noteworthy that whereas sPMv was activated quite frequently by passive listening to syllables or words, it was not activated in any study of sentence comprehension, and it was activated in only two studies of narrative comprehension. This appears to be paradoxical, because presumably the phonetic, phonological and lexical access processes involved in syllable and word comprehension are also necessary for sentence and narrative comprehension.

To resolve this paradox, we propose that a role of sPMv is to direct attention to the *phonetic form* of auditory stimuli. When sentences and narratives are perceived, there is little attention to the phonetic form per se; rather, we simply perceive the meaning of the language that we hear. On the other hand, when syllables are presented in isolation in an experimental context, they have no meaning, and if the subject is to attend to the stimuli at all they will likely attend to the actual physical form of the stimuli. Words are an intermediate case since they have meanings which could be accessed. However it seems plausible that when isolated words are presented in meaningless sequences in the scanner, subjects may frequently attend less to word meaning than they do to the sound of the words. The proposed role of sPMv in attention is bolstered by several studies of auditory attention, which have reported activations in similar regions (Tzourio et al., 1997; Zatorre et al., 1999; Lipschutz et al., 2002; Degerman et al., 2006). Activations in these studies involving non-linguistic stimuli are typically strongly right-lateralized.

It is possible that sPMv is the generator of a frontal component of the N100 auditory response and/or of the mismatch negativity (MMN). The N100 is a negative wave peaking about 100 ms after the onset of an auditory stimulus. Näätänen & Picton (1987) argued that there are at least three components to the N100, the third of which they hypothesized to be generated by motor and premotor cortex, and responsible for transient arousal facilitating motor activity. The mismatch negativity (MMN), which is elicited by discriminable auditory changes in a train of standard stimuli, peaks somewhat later than the N100 but overlaps with it in terms of latency. The MMN is also believed to have a frontal component, which may reflect a mechanism to direct attention to novel stimuli detected in temporal areas (Giard et al., 1990). This theory suggests that the frontal component would occur later in time than the temporal component, which has recently been confirmed (Rinne et al., 2000). A previous PET study of auditory attention observed precentral activity and argued that it may reflect N100 or MMN generators (Tzourio et al., 1997). Recent electrophysiological studies involving direct recordings in awake neurosurgical patients have demonstrated that a region in the vicinity of sPMv is engaged very rapidly in response to auditory stimulation (Edwards, 2006; see also Edwards et al., 2005). Edwards (2006) and colleagues observed high gamma activity around the precentral gyrus and central sulcus when subjects were presented with auditory words. This activity was clearly distinct from background noise well before 100 ms, and peaked at about 300 ms. It is possible that the early phases of this activity might constitute frontal generators for the N100 and/or the MMN. Note that in the earlier study, Edwards et al. (2005) did not find evidence for frontal involvement in the MMN, but most electrodes were located more ventrally and anteriorly in the IFG. Other neuroimaging (Opitz et al., 2002; Doeller et al., 2003; Rinne et al., 2005; Yucel et al., 2005) and MEG (Pulvermüller et al., 2003) studies have suggested a more inferior locus for frontal MMN components.

A final noteworthy finding is that a region in the vicinity of sPMv was activated by viewing of single letters (Longcamp et al., 2003); the location of this activation is shown in Figures 1.2 and 1.4. Longcamp et al. (2003) found that there was overlap between activation for reading single letters and writing them, although the two peaks for writing did not correspond exactly to the peak for reading: one was more dorsal, probably in PMd, whereas the other was more ventral ($z = 29$). Longcamp et al. (2005) found that whereas the reading activation was left-lateralized in right-handed subjects, it was right-lateralized in left-handed subjects. These observations are consistent with the proposal above that sPMv is involved in attention to phonetic form. The reading and writing of letters in a non-meaningful context may involve attention to their phonetic values even though no auditory stimulus is physically present.

Finally a third frontal region was frequently activated by perception of words, sentences and narratives: a ventral area around the border of BA 44 and BA 45. This region was not activated in speech production studies, and hence these activations probably do not reflect motor system involvement in speech perception. Rather they probably reflect lexical access (semantics) and/or syntactic processing, which are functions of the more anterior parts of the IFG (Dapretto & Bookheimer, 1999; Poldrack et al., 1999; Bookheimer, 2002).

1.7 Summary and conclusions

In sum, there is compelling evidence that frontal speech motor regions play a role in overt phonological processing and phonetic discrimination and identification tasks. Classic phonological processing studies (e.g. Zatorre, 1992, 1996) showed that left frontal regions in the vicinity of dorsal BA 44 were activated for phonological processing. Studies of aphasia revealed deficits in phonetic discrimination and identification following frontal lesions (Gainotti et al., 1982), and demonstrated relationships between phonemic output deficits and phonetic discrimination deficits (Basso et al., 1977; Miceli et al., 1980). Cortical excision of face motor cortex was shown to profoundly impair phonetic identification (Taylor, 1979). Direct electrical stimulation of cortex showed that the same sites which disrupt orofacial mimicry also frequently impair phonetic identification (Ojemann & Mateer, 1979; Ojemann, 1981, 1983). A recent rTMS showed that temporary inactivation of sPMv disrupted performance in a phonetic identification task (Meister et al., submitted). However the extent to which frontal regions are involved in speech perception under normal circumstances is controversial, with some researchers arguing that the auditory-motor pathway is important primarily for speech production and language acquisition (Hickok & Poeppel, 2000, 2004), while others have contended that the motor system may be involved in speech perception more generally (Fadiga et al., 2002; Watkins et al., 2003; Wilson et al., 2004; Pulvermüller et al., 2006), perhaps especially in situations where the acoustic input is degraded in some manner (Callan et al., 2004).

The strongest evidence in support of the first view comes from the relatively good comprehension performance of patients with frontal lesions, and the dissociations between phonetic tasks and comprehension frequently observed in patients (e.g. Blumstein et al., 1977a). Relatedly, Coleman (1998) argued that phonological entries in the mental lexicon are auditory rather than motor, based in part on the fact that frontal areas were rarely activated in early studies of speech perception. As discussed above, more recent studies have frequently reported motor activity during speech perception, but it remains possible that phonological forms in the lexicon are basically auditory, with the motor system involved in their access but not in actual representations. In favor of the second view—that the motor system is involved in normal speech perception, perhaps especially in perceptually challenging situations—is the finding that the performance of Broca's aphasics on lexical comprehension was differentially affected under noise (Moineau et al., 2005), and several imaging studies which have reported activation in frontal regions in conditions requiring effortful phonetic processing (e.g. Poldrack et al., 2001; Davis & Johnsrude, 2003; Giraud et al., 2004). We favor the view that the motor system is involved in everyday speech perception; it seems implausible that frontal regions could be so indispensable for performing phonetic and phonological tasks if they were not accustomed to playing a perceptual role in everyday situations.

We identified two major candidate premotor regions which may play a role in speech perception. One is dorsal BA 44, which is frequently activated in phonological processing studies (e.g. Zatorre et al., 1996), under challenging phonetic conditions (e.g. Davis & Johnsrude, 2003), and by passive listening. Many speech production studies have reported peaks in this vicinity, and electrical stimulation and TMS confirm that this is a speech premotor region. The other is sPMv in BA 6, which is frequently activated by passive listening to syllables or words (e.g. Binder et al., 2000; Wilson et al., 2004). Stimulation to this region impairs phonetic identification (Meister et al., submitted) confirming its essential role in speech perception tasks. We have argued that sPMv is important for attention to phonetic form and along with dorsal BA 44 constitutes a crucial component of the dorsal auditory-motor stream.