

The neural correlates of inner speech defined by voxel-based lesion–symptom mapping

Sharon Geva,¹ P. Simon Jones,¹ Jenny T. Crinion,² Cathy J. Price,³ Jean-Claude Baron^{1,4} and Elizabeth A. Warburton¹

1 Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK

2 Institute of cognitive Neuroscience, University College London, London WC1N 3AR, UK

3 Wellcome Trust Centre for NeuroImaging, Institute of Neurology, London WC1N 3BG, UK

4 INSERM UMR 894, Université Paris 5, Sorbonne Paris Cité, 2 ter Rue d'Alésia, 75014 Paris, France

Correspondence to: Sharon Geva,
Department of Clinical Neurosciences,
University of Cambridge,
R3 Neurosciences,
Box 83, Addenbrooke's Hospital,
Cambridge CB2 0QQ, UK
E-mail: sharon.geva@cantab.net

The neural correlates of inner speech have been investigated previously using functional imaging. However, methodological and other limitations have so far precluded a clear description of the neural anatomy of inner speech and its relation to overt speech. Specifically, studies that examine only inner speech often fail to control for subjects' behaviour in the scanner and therefore cannot determine the relation between inner and overt speech. Functional imaging studies comparing inner and overt speech have not produced replicable results and some have similar methodological caveats as studies looking only at inner speech. Lesion analysis can avoid the methodological pitfalls associated with using inner and overt speech in functional imaging studies, while at the same time providing important data about the neural correlates essential for the specific function. Despite its advantages, a study of the neural correlates of inner speech using lesion analysis has not been carried out before. In this study, 17 patients with chronic post-stroke aphasia performed inner speech tasks (rhyme and homophone judgements), and overt speech tasks (reading aloud). The relationship between brain structure and language ability was studied using voxel-based lesion–symptom mapping. This showed that inner speech abilities were affected by lesions to the left pars opercularis in the inferior frontal gyrus and to the white matter adjacent to the left supramarginal gyrus, over and above overt speech production and working memory. These results suggest that inner speech cannot be assumed to be simply overt speech without a motor component. It also suggests that the use of overt speech to understand inner speech and vice versa might result in misleading conclusions, both in imaging studies and clinical practice.

Keywords: stroke; aphasia; inner speech; voxel-based lesion–symptom mapping

Introduction

Inner speech, or the ability to speak silently in one's head, has been suggested to play an important role in memory (Baddeley and Hitch, 1974), reading (Corcoran, 1966), language acquisition

(Vygotsky, 1962), language comprehension (Blonskii, 1964), thinking (Sokolov, 1972) and even in consciousness and self-reflective activities (Morin and Michaud, 2007).

Currently, two main levels of inner speech may be differentiated from the available literature: The first level is abstract inner speech

Received January 20, 2011. Revised August 8, 2011. Accepted August 10, 2011

© The Author (2011). Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

or 'the language of the mind'. The first to investigate it using the methodology of experimental psychology were Egger (1881) and Ballet (1886). By using introspection, they tried to understand the relation between inner speech and thought and by doing so they also brought about an outburst of experimental work on inner speech (reviewed in Sokolov, 1972). Later, Vygotsky (1962) argued that young children have no inner speech and therefore they can only think out loud. With the acquisition of language, speech becomes increasingly internalized. Mature inner speech, he argued, is different from overt speech in that it lacks the complete syntactic structures available in overt speech, and its semantics is personal and contextual rather than objective.

The second level is of concrete inner speech. It is flexible and can therefore be either phonological or phonetic (Oppenheim and Dell, 2010; see Vigliocco and Hartsuiker, 2002 for a related distinction). Phonological inner speech displays the 'lexical bias effect' (the tendency for errors in speech production to produce other words rather than non-words) but not the 'phonemic similarity effect' (the tendency to mix similar phonemes in speech production), suggesting that it is phonetically impoverished in comparison to overt speech (Oppenheim and Dell, 2008). Phonetic inner speech, on the other hand, displays both types of biases (Oppenheim and Dell, 2010). Ozdemir *et al.* (2007) examined the influence of the 'uniqueness point' of a word on monitoring for the presence of specific phonemes in a word. A word's uniqueness point is the place in its sequence of phonemes at which it deviates from all other words in the language; hence, it makes the word 'unique'. They reported that 'uniqueness point' influenced inner speech, therefore suggesting that its phonetic components are similar to that of overt speech. In a study that looked at inner speech monitoring, participants were asked to produce 'tongue twisters' and report the number of self-corrections (Postma and Noordanus, 1996). Participants repeated the task in different conditions: inner speech, mouthing, overt speech in the presence of white noise and overt speech without noise. Interestingly, there was no difference in the number of errors detected by the participant in the first three conditions. Together, these two studies also give evidence to the existence of a phonetically rich inner speech. In this study, we investigated concrete inner speech. Inner speech was defined as the ability to create an internal representation of the auditory word form, and to apply computations or manipulations to this representation.

Patients with post-stroke aphasia often complain that there is poor correspondence between the words they think or intend to say (inner speech), and the words they are able to produce out loud (overt speech) (Marshall *et al.*, 1994). Indeed, there is some evidence showing that inner and overt speech can dissociate in aphasia. Feinberg *et al.* (1986) tested five patients with conduction aphasia who were unable to read words aloud. Four of the five demonstrated intact performance on inner speech tasks such as judgement of word length, and judgement of whether pairs of words were homophones or rhymes, all using pictures. Marshall *et al.* (1985) presented a case study of a patient who had severe auditory comprehension deficits and impairment in speech production. Despite this, she often corrected her own errors and was relatively successful on various inner speech tasks, including rhyme and homophone judgement and phoneme monitoring in

reading. Recently, we studied a group of 27 patients with chronic post-stroke aphasia, using tests for language abilities, speech apraxia and inner speech (homophone and rhyme judgements, using both words and pictures). We have shown that while for most patients with aphasia there is a high correlation between inner and overt speech abilities, some show preserved inner speech together with a marked deficit in overt speech, while others show the reverse pattern: impaired inner speech together with relatively intact overt speech. These results suggest that inner speech can be, at least in some cases, dissociated from overt speech, and that inner speech is dependent on both the production and the comprehension systems (Geva *et al.*, 2011).

Another source of information regarding the differences between inner and overt speech comes from brain imaging studies of language in normal subjects. Many of these studies use a covert response (inner speech) as the preferable response mode, apparently assuming that overt and inner speech differ only in the articulatory motor component present in overt speech. However, other studies run contrary to this assumption (Huang *et al.*, 2002; Gracco *et al.*, 2005; Shuster and Lemieux, 2005). Direct comparisons between conditions of overt and inner speech indicate that although they yield overlapping brain activation, the two conditions also produce separate activations in other regions of the brain, reflecting distinct non-motor cognitive processes (Ryding *et al.*, 1996; Barch *et al.*, 1999; Palmer *et al.*, 2001; Huang *et al.*, 2002; Indefrey and Levelt, 2004; Shuster and Lemieux, 2005; Basho *et al.*, 2007).

When studying inner speech using functional imaging, participants are asked to covertly perform tasks such as semantic or phonological fluency, verb generation or stem completion, among others. In these cases, the experimenter cannot reliably determine whether participants perform the task using the desired cognitive processes or whether they perform the task at all. If the task is performed, in some cases it might be that 'lower' levels of inner speech are used, such as the abstract or phonetically impoverished ones, and the researchers cannot distinguish between, or control these cases. Additionally, informative and important data regarding performance (type of response, errors and reaction time) cannot be obtained (Barch *et al.*, 1999; Peck *et al.*, 2004). Lastly, some studies do not ensure that participants refrain from producing overt speech when asked to generate only inner speech (reviewed in Indefrey and Levelt, 2004).

Rhyme judgement has been used previously in imaging studies (Frith, and Frackowiak, 1993; Pugh *et al.*, 1996; Lurito *et al.*, 2000; Paulesu *et al.*, 2001; Hoeft *et al.*, 2007), and unlike the tasks mentioned above, can provide the experimenter with control over, and data regarding, subjects' performance. Hoeft *et al.* (2007) found that covert rhyme judgement created significant activation in the left hemisphere, in posterior parts of the middle and inferior frontal gyrus, the inferior parietal lobule and lateral regions of the occipital lobe, extending into the inferior temporal lobe. In another study, covert word rhyme judgement task was compared with a baseline task in which participants performed a similarity judgement on sets of lines (Lurito *et al.*, 2000). Activation was found in the left hemisphere in the middle frontal gyrus, inferior frontal gyrus, supramarginal gyrus, middle temporal gyrus and fusiform gyrus. In the right hemisphere, activation was found in the

inferior frontal gyrus. Paulesu *et al.* (1993) found that covert rhyme judgement activated Brodmann area (BA) 44 as well as motor regions. They suggest that while BA 44 is essential for phonological processing, activation in the motor regions is probably related to small laryngeal movements that are not essential for the production of inner speech. Comparing a covert rhyme judgement task to a letter case judgement task, activation was found only in Broca's area (BA 44/45) (Poldrack *et al.*, 2001). Lastly, Pugh *et al.* (1996) investigated activation associated with non-word covert rhyme judgement by using a region of interest analysis that included the lateral orbital gyrus (BA 10 and 47), dorsolateral prefrontal cortex (BA 46), inferior frontal gyrus (BA 44 and part of 45), superior temporal gyrus (BA 22, 38 and 42), middle temporal gyrus (BA 21, 37 and 39), lateral extrastriate cortex (BA 18 and 19) and medial extrastriate cortex. The study showed that the frontal areas (lateral orbital gyrus, dorsolateral prefrontal cortex and inferior frontal gyrus) are specific to phonological processing, while the temporal areas (superior and middle temporal gyri) are involved in phonological processing among other functions.

In summary, in previous functional imaging studies the covert rhyme judgement task most commonly activated the left inferior frontal gyrus. Two of the studies also found activation in inferior parietal regions. These studies provide vital information for understanding inner speech, but in order to understand how inner speech differs from overt speech, a direct comparison between the two must be made.

Comparing overt and covert responses directly, Basho *et al.* (2007) found that a covert response to a semantic fluency task produced significantly greater activation in the left middle temporal gyrus (BA 21), left superior frontal gyrus (BA 6), right cingulate gyrus (BA 32), right superior frontal gyrus (BA 11), right inferior and superior parietal lobe (BA40 and 7) and the left parahippocampal gyrus (BA 35/36). The authors attribute some of this activation to inhibition of overt response and response conflict (producing a word but not saying it aloud). Inner speech produced during a word repetition task showed higher activation than overt speech in the left middle frontal gyrus and paracentral lobule, as well as in some right hemispheric regions including the postcentral gyrus, two regions in the middle temporal gyrus, the precuneus and the cerebellum (Shuster and Lemieux, 2005). Huang *et al.* (2002) conducted a region of interest analysis, looking at the mouth areas of the primary motor cortex, an area just inferior to it, and Broca's area. They found that Broca's area showed a task-dependent pattern of activation. While in a letter naming task, activation was greater for overt speech, in a task requiring generating animal names, activation was greater for inner speech. The authors suggest that increase of activation during silent speech is related to either phonological processing or to the inhibition of an overt response. It is unclear, however, why phonological processing should differ between inner and overt speech and the issue is not addressed in the study. Similarly, Bookheimer *et al.* (1995) found that silent reading, during a PET scan, showed increased activation in the left inferior frontal gyrus, compared with reading aloud. Comparing a covert rhyme judgement task with an overt homophone reading task, significant activation was found in left precentral gyrus (BA 6), left supramarginal gyrus (BA 40,

bordering BA 39), left inferior parietal lobe (BA 40) and left dorsal frontal cortex (Owen *et al.*, 2004). Focusing on cerebellar activation, Frings *et al.* (2006) found that silent verb generation was associated with greater activation in a specific right cerebellar region, when compared with overt reading aloud of the same verbs. Non-cerebellar activations were found in the left superior temporal gyrus and in the left inferior frontal gyrus. It is difficult to determine, however, whether the observed activation is related to the difference between inner and overt speech or to differences in task demands (i.e. verb generation versus reading aloud). All but one of the studies above (Basho *et al.*, 2007) found greater activation in the overt condition in various brain regions including motor and pre-motor regions related to articulation (Huang *et al.*, 2002; Owen *et al.*, 2004; Shuster and Lemieux, 2005; Frings *et al.*, 2006); sensory area (Shuster and Lemieux, 2005; Frings *et al.*, 2006); and other regions [superior temporal sulcus (Bookheimer *et al.*, 1995; Shuster and Lemieux, 2005); supramarginal gyrus (Bookheimer *et al.*, 1995); right inferior occipital gyrus (Owen *et al.*, 2004); and left inferior frontal gyrus including BA 44 and 45 (Owen *et al.*, 2004)].

In summary, these studies show that a number of regions in the brain are activated during various inner speech tasks, when compared with overt speech production. This occurs in both hemispheres as well as in the cerebellum and the results of the various studies diverge significantly. It is important to note that some of these studies also have several of the caveats mentioned above, namely, failing to control for performance on the inner speech condition and not ensuring that participants refrain from producing overt speech.

In conclusion, studies of inner speech alone produce replicable data regarding inner speech but in those studies the relation between inner and overt speech is not explored. Other studies reviewed here made direct comparison between inner and overt speech but used tasks that do not monitor participants' performance. The purpose of the current study was to further our understanding of the neural mechanisms underlying inner speech and its relation to overt speech, while controlling for participants' performance. In aphasia, lesion analysis together with detailed behavioural testing can give information regarding the neural correlates of inner speech which cannot be easily obtained by functional MRI. Lesion analysis can define the areas that are critical for, rather than only contributing to, the production of inner speech. Lesion analysis also avoids the difficulties in using overt and covert speech in functional imaging studies, discussed previously.

In this study, the anatomical correlates of inner speech and their relation to overt speech and working memory were examined, using voxel-based lesion–symptom mapping (Bates *et al.*, 2003). Rhyme and homophone judgement tasks were used to assess inner speech, a reading aloud task was used to assess overt speech and a sentence repetition task was used to assess verbal working memory. A well-documented finding shows that rhyme judgement requires working memory, while homophone judgement does not (reviewed in Howard and Franklin, 1990). A common way of testing whether the performance of a cognitive task requires the resources of the working memory system is by examining what types of additional tasks or stimuli interfere with performance. The 'articulatory control process' is the part of the working memory

system that allows refreshing the memory trace of the information inside the phonological store by using subvocal rehearsal. Its normal function can be disrupted by the recitation of irrelevant material, a phenomenon known as 'articulatory suppression'. It has been shown that when subjects are asked to recite irrelevant material aloud, their performance on rhyme judgement, but not on homophone judgement, declines (Kleiman, 1975; Besner *et al.*, 1981; Wilding and White, 1985; Johnston and McDermott, 1986; Brown, 1987; Richardson, 1987; Howard and Franklin, 1990). These data have been taken by many as evidence to support the idea that rhyme judgement requires working memory, while homophone judgement does not. Therefore, in the analysis of the rhyme judgement task, working memory scores were included as well. Based on previous studies, it was hypothesized that one or both of the most commonly activated regions, the left inferior frontal gyrus and left inferior parietal lobe, will prove to be essential for inner speech.

Materials and methods

Participants

Twenty-one patients participated in the study (14 males/7 females; age range: 21–81 years; mean age: 64 ± 15 years; mean number of years of education: 12 ± 3 ; mean time since last stroke: 27 ± 21 months). The inclusion criteria were as follows: the development of aphasia following a left middle cerebral artery territory stroke, 18 years of age or above, native speakers of English, no history of neurological or psychiatric disorders other than stroke and no major cognitive impairment. The diagnosis of stroke was made by clinicians who saw the patient on admission, according to the clinical definition: acute onset of focal symptoms persisting for >24 h. Both acute and follow up imaging (CT and MRI) confirmed the diagnosis. Patients with transient ischaemic attack but no diagnosis of stroke were excluded from the study. Two of the patients who had more than one stroke had no signs of the first stroke on MRI, when scanned clinically after that stroke. The diagnosis of aphasia was based on the convergence of clinical consensus and the results of a standardized aphasia examination, the Comprehensive Aphasia Test (Swinburn *et al.*, 2004). Patients had impaired production of speech but relatively preserved comprehension (to a level allowing them to consent to the study and understand the behavioural tasks). To exclude the possibility that patients have other major cognitive impairments, patients were also given a set of cognitive tests, including the Brixton Test of executive functions (Burgess and Shallice, 1997), the Rey-Osterrieth Complex Figure Test (Meyers and Meyers, 1995) and parts of the Addenbrooke's Cognitive Examination - Revised (ACE-R), testing visual-spatial abilities (Mathuranath *et al.*, 2000). All patients performed above the standardized cut-off scores for these tests. Table 1 presents additional demographic and clinical information for patients who completed the entire study. The study was approved by the Cambridge Research Ethics Committee and all participants read an information sheet and gave written consent.

Behavioural testing

The inner speech tests were adapted from the Psycholinguistic Assessments of Language Processing in Aphasia (Kay *et al.*, 1992). In the first task, participants were asked to determine whether two

written words rhyme. For example, 'bear' and 'chair' rhyme, while 'food' and 'blood' do not. The test had a total of 60 pairs. Half the rhyming pairs and half the non-rhyming pairs had orthographically similar endings (e.g. town–gown versus hush–bush), while the other half had orthographically dissimilar endings (e.g. chair–bear versus bond–hand). This allowed us to determine whether patients are using their inner speech or resorting to an alternative cognitive strategy, in this case, solving the task using orthography. A patient who solves the task based on orthography alone will score 50% correct, which is chance level. In the second task, participants had to determine whether two words sound the same, i.e. whether they are homophones. This test had 40 pairs. For example, 'might' and 'mite' are homophones, while 'ear' and 'oar' are not. The tasks could not be successfully solved based on orthography alone, therefore ensuring that the participants had to use their 'inner speech' to solve the tasks. Prior to the test, patients were given instructions and then a practice that included a minimum of 10 items (more, when needed). In the practice, the experimenter read the items aloud first and the patient had to give his/her judgement. This was done to confirm that the patient understood the task and had no significant receptive phonological impairment. The criterion employed was that the patient had to answer five consecutive trials correctly. For almost all patients this was achieved in the first five trials. In cases where patients made errors in the first five trials, more trials were given until the criterion was met. If the criterion was not met after a maximum of 20 trials, the task was discarded. The practice was also used to make sure that patients solve the task without using overt speech; patients practiced the task until they were able to perform the tasks without producing any sound or articulatory movements.

Patients marked their answers using an answer sheet. On the sheet, two columns were presented; one with the symbol ✓ and the written word 'YES', and the other with the symbol ✗ and the written word 'NO'. Patients were asked to mark their answer in the correct column. In cases where the patient found this difficult, he/she pointed at their answer (✓ YES or ✗ NO) and the examiner marked the answer on the answer sheet.

The words in each task were randomly assigned to one of two lists. Participants performed the task on half of the items (one list) using inner speech and half (second list) using overt speech. In both cases, the patient would first read both words in the pair (either internally or overtly, depending on the condition) and then give his/her judgement for the pair. The two conditions were completed separately and successively, and the order of conditions was randomized between patients. Patients who were unable to read aloud performed the entire task (both lists) using inner speech alone. These patients were defined as those who scored less than one-third correct on the word reading aloud task in the Comprehensive Aphasia Test. Scoring of the inner speech task was based on the judgement given to a word pair, with possible answers being correct or incorrect. Hence, every pair judged incorrectly was scored as one error.

Overt speech was scored as follows: two points were given when a word was read correctly and one point when the word was initially read incorrectly but a self-correction subsequently took place without prompting. No points were given for words read incorrectly.

Some patients with aphasia demonstrate perseverations and automatic speech in overt speech. In order to make the overt and covert conditions as equivalent as possible, we scored perseverations or automatic speech as an incorrect response in the overt speech task. This way, if these deficits influence inner speech as well the scores of the inner speech task will be comparable with regard to this aspect, with those of the overt speech task.

Table 1 Demographic and clinical information and performance on the Comprehensive Aphasia Test and the Apraxia Battery for Adults

Patient	Age	Sex	Time since stroke (months)		Stroke type	Handedness ^c	Scan	Auditory comprehension of words	Reading comprehension of words	Object naming	Word repetition	Speech apraxia
			Previous stroke ^a	Last stroke ^b								
1	66	M		10	Ischaemic	R	MRI (3T)	28	30	18	12	Moderate
2	69	M		17	Ischaemic	R	MRI (3T)	26	28	0	0	Mild
3	73	M	29 (RH, ischaemic)	16	Ischaemic	L	MRI (3T)	29	29	43	29	None
4	62	M		10	Haemorrhagic	R	MRI (3T)	30	30	48	32	None
5	78	M	92 (RH, ischaemic)	64	Ischaemic	A (−0.1)	MRI (3T)	30	26	14	6	Moderate
6	69	M	Transient ischaemic attack	25	Ischaemic	R	MRI (3T)	27	30	46	30	None
7	78	F		9	Ischaemic	R	MRI (3T)	28	29	45	32	None
8	78	M	50 (LH, ischaemic)	12	Ischaemic	L	MRI (3T)	24	22	24	25	None
9	21	F		15	Ischaemic	R	MRI (1.5T)	24	28	40	32	Mild
10	42	F		12	Ischaemic	R	MRI (3T)	30	30	40	32	Severe
11	81	M	72 (LH, ischaemic)	19	Ischaemic	R	MRI (3T)	29	24	13	26	None
12	62	M		28	Ischaemic	R	MRI (3T)	30	30	45	30	None
13	65	F		24	Haemorrhagic	R	MRI (1.5T)	29	30	34	18	Mild
14	71	M		60	Ischaemic	R	MRI (1.5T)	21	12	11	7	Severe
15	79	M	120 (RH, ischaemic)	8	Ischaemic	L	MRI (3T)	29	30	46	32	None
16	49	F		20	Ischaemic	R	MRI (3T)	28	30	46	32	None
17	53	F		24	Ischaemic	R	MRI (3T)	23	22	15	29	Severe

a For patients who had more than one stroke, time since the first stroke is indicated.

b First stroke for patients who had only one, second for those patients who had a previous stroke. Last stroke was left hemispheric in all cases and the cause of the language deficits.

c In brackets: the score received on the Edinburgh Handedness Inventory for ambidextrous subjects, where -1 = strongly left handed, 1 = strongly right handed and 0 = completely ambidextrous.

A = ambidextrous; L = left; LH = left hemisphere; R = right; RH = right hemisphere.

Verbal working memory was tested using a spoken sentence repetition task that had sentences varying in length (from three to six content words, two sentences of each kind). Subjects heard a sentence spoken by the examiner and were asked to repeat it. The score given corresponded to twice the number of content words in the longest sentence repeated successfully. Errors that were recognizable versions of words in the target sentence (minor phonemic errors, apraxic errors and dysarthric distortions) were accepted as correct. Perseverations and automatic speech were not scored as incorrect. In such cases, the patient was given another opportunity to repeat the sentence. This scoring system allowed the task to reflect sentence span only, with minimal influence from other production deficits.

Other tasks involving single word production and comprehension (taken from the Comprehensive Aphasia Test; Swinburn *et al.*, 2004), include: (i) auditory comprehension of words: the examiner read aloud a word to the subject, who was asked to point to one of four pictures that best goes with the word. The task had 15 trials; (ii) reading comprehension of words: this task was identical to the previous one, only this time the word was written in the middle of the page, instead of spoken out loud by the examiner; (iii) word repetition: subjects were asked to repeat words read out loud by the examiner. This task included 16 short words; (iv) object naming: subjects were asked to name 24 pictures of nouns. In all four tasks, a correct answer was given 2 points. A delayed answer or a correct answer following

self-correction was given 1 point. In the auditory word comprehension and the word repetition tasks, if the participant asked the examiner to repeat the question, and this was followed by a correct answer, 1 point was given as well; and (v) speech apraxia: subjects performed two subtests from the Apraxia Battery for Adults (Dabul, 1979): in one they were asked to repeat various combinations of syllables read out by the examiner; and in the second, limb and oral apraxia were examined by asking subjects to perform various hand and oral motor actions. Severity of apraxia was defined as: 0–29 severe; 30–39 moderate; 40–49 mild; 50 none. Scores for all tasks are reported in Table 1.

Imaging data acquisition

Imaging was performed using a 3T Siemens Allegra MRI scanner. Four patients could not undergo a 3T MRI scan due to cardiac stents ($n = 2$) or a patent foramen ovale device ($n = 2$), which were not 3T compatible. These patients were scanned using a 1.5T MRI Siemens scanner. Imaging included proton density and T_2 -weighted scans (repetition time: 4.6 s, echo time: 12 ms for proton density, 104 ms for T_2 , field of view: 168×224 mm, matrix: 240×320 , sagittal plane; slice thickness: 5 mm; 27 slices), a magnetization-prepared rapid acquisition gradient echo (MPRAGE) scan (repetition time: 2.3 s, echo time: 2.98 ms, field of view: 240×256 mm, sagittal plane; slice thickness: 1 mm; 176 slices) and an axial fluid-attenuated inversion recovery (FLAIR) scan

(repetition time: 7.84 s, echo time: 95 ms, field of view: 256 × 320 mm, axial plane; slice thickness: 4 mm; 27 slices).

Data analysis

Lesions were defined using the region of interest facility in Analyze software (Mayo Biomedical Imaging Resource, Mayo Clinic). One author (S.G.) drew the lesions manually on T₂ scans, while consulting the other sequences. The drawn lesions were validated by a trained neurologist (E.A.W.) who was blinded to the patients' diagnosis. Masks were made from the lesions using MRIcron (Chris Rorden, MRIcron 2009) and these were used as inclusive masks in the spatial segmentation routine of the Statistical Parametric Mapping software (SPM5, Wellcome Department of Cognitive Neurology) implemented in the MATLAB (2006b, version 7, The MathWorks Inc.) environment. The spatial parameter files were then applied to the original drawn lesion which resulted in a spatially normalized binary lesion definition for each patient. A lesion overlap map is shown in Fig. 1.

For statistical analysis, only voxels in which at least 20% of the patients had a lesion were included in the analysis. Patients with multiple strokes were not excluded, because the voxel-based analysis used in voxel-based lesion–symptom mapping does not take into account how the effect of damage to one voxel depends on the effect of damage to other voxels. Rather, it looks for the most significant correlations between behaviour and damage, irrespective of damage to other brain regions. Effective coverage map defined the regions where effects could and could not possibly be detected at a given significance threshold of $\alpha = 0.05$. The maps were calculated based on the number of patients who have a lesion in each voxel and their distribution of behavioural scores (Fig. 2) (for further details and an example see Rudrauf *et al.*, 2008, p. 10). Figure 2 represents effective coverage map for the rhyme judgement task. Due to the high correlation between the behavioural scores of the homophone and rhyme judgement tasks, the effective coverage maps were very similar and hence only the map for the rhyme judgement task is presented. As can be seen in Fig. 2, there was sufficient effective coverage in the region of

the inferior frontal gyrus and inferior parietal lobe, which were the main focus of interest in the current study, as well as in the superior temporal gyrus and temporal pole, insula and the precentral gyrus.

Statistical analysis was done using voxel-based lesion–symptom mapping (Bates *et al.*, 2003). In voxel-based lesion–symptom mapping, patients are divided into two groups according to whether they do or do not have a lesion affecting a specific voxel. Behavioural scores are then compared for these two groups, yielding a *t*-statistic for that voxel. The procedure is then repeated several times for each voxel included in the analysis. The covariate of interest (rhyme or homophone judgment) was first examined by itself using the NPM (non-parametric mapping) software package (Rorden *et al.*, 2007). A *t*-statistic was calculated and correction for multiple comparisons was achieved by employing the non-parametric permutation test, as recommended for medium-sized samples (Kimberg *et al.*, 2007; Medina *et al.*, 2010). Data were permuted 1000 times with each permutation resulting in a calculated *t*-statistic. The distribution of those *t*-statistics was used to determine the cut-off score at $P < 0.05$.

After identifying the areas that are significantly associated with the homophone and rhyme judgement, we carried out analyses aimed at isolating specific cognitive components (see below), or examining the influence of other variables on the data. Both covariates were entered into a single multiple regression analysis in voxel-based lesion–symptom mapping version 1.6 (Bates *et al.*, 2003), and correction for multiple comparisons was done using false discovery rate, at a threshold of $P < 0.05$. Voxel-based lesion–symptom mapping maps were created based on these cut-off scores, and images show these statistical maps overlaid onto the template provided in voxel-based lesion–symptom mapping.

Analysis of cognitive components

To distinguish the different components of each of the inner speech tasks, measurements of overt speech production and verbal working memory were included. In the inner speech tasks, the following cognitive processes take place: visual word processing, grapheme to phoneme translation, inner speech, and in the rhyme judgement, verbal

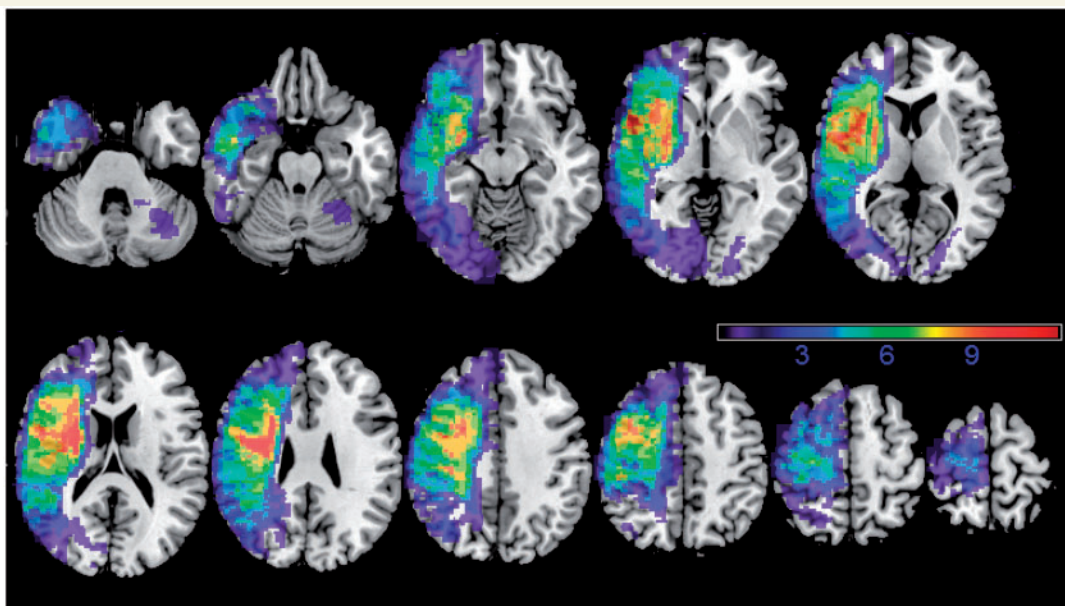


Figure 1 An overlay of all patients' lesions. Warmer areas indicate areas of greater lesion overlap.

working memory. Phonetic coding and subvocal articulation can accompany inner speech production, but they are not a necessary component of it. Table 2 specifies the cognitive processes involved in each task.

The comparisons of interest in this study were as follows: (i) homophone judgement covaried for reading words aloud, which examines inner speech alone; (ii) rhyme judgement covaried for reading words aloud, which examines inner speech and working memory; and (iii) rhyme judgement covaried for sentence repetition, which removes the working memory component in the rhyme judgement task.

Other variables used as covariates in the voxel-based lesion–symptom mapping analyses included time post-stroke (in months), age and lesion volume. To evaluate the influence of handedness, all analyses were done with and without the left-handed patients.

Results

Behavioural results

Twenty patients completed the homophone judgement task and 18 of them also completed the rhyme judgement task. Detailed behavioural results were reported elsewhere (Geva *et al.*, 2011).

In short, the average score on the rhyming task was $76 \pm 17\%$ (range: 47–100%). The average score on the homophone task was $77 \pm 21\%$ (range: 42–100%). Correlations between age, time since stroke or lesion volume and performance on the inner speech tasks were tested using Kendall's Tau (significance level was determined at $P < 0.005$, after a Bonferroni correction for multiple comparisons). The correlation between the two inner speech tasks was significant (Kendall's Tau = 0.68, $P < 0.001$). Age and time since stroke did not significantly correlate with performance on the inner speech tasks ($P > 0.005$ for all). Larger lesions were significantly correlated with poorer performance on the rhyming task (Kendall's Tau = -0.68 , $P < 0.001$) and on the homophone task (Kendall's Tau = -0.49 , $P = 0.004$). Table 1 shows each patient's scores on the language tasks taken from the Comprehensives Aphasia Test and the scores of the Apraxia Battery for Adults.

Voxel-based lesion–symptom mapping results

There were technical problems with the scan of one patient (D.R., scanned at 1.5T) and the data of this patient were excluded from

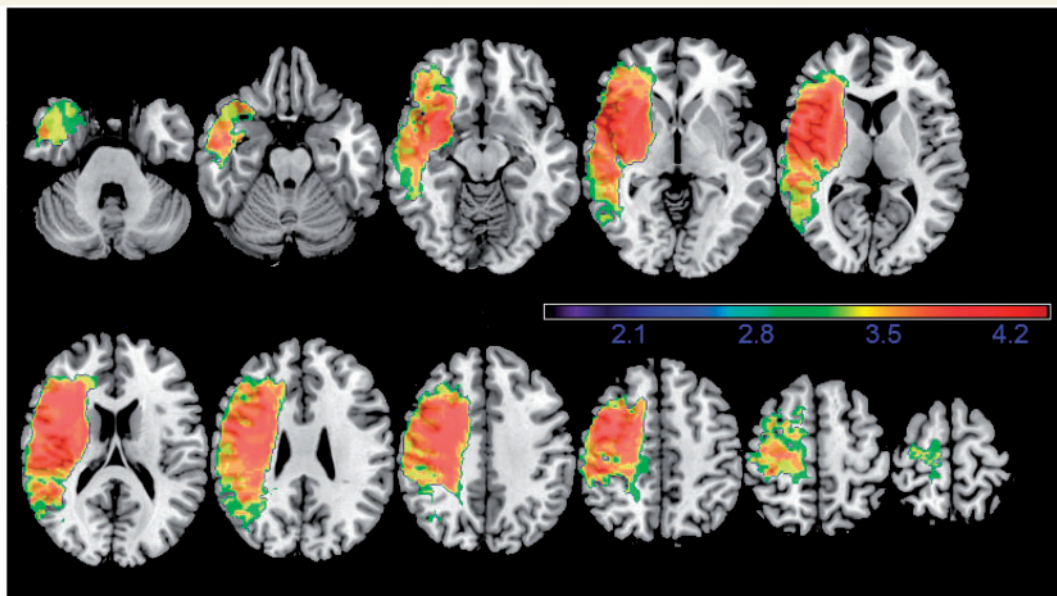


Figure 2 Map showing distribution of effective coverage for the rhyme judgement task for voxels in which at least 20% of patients had a lesion, thresholded at $P < 0.05$. Warmer colours represent higher power. Colours represent Z-scores, running from 1.64 to the highest Z-score in the image.

Table 2 Cognitive subprocesses involved in the inner speech tasks

Test	Visual word processing	Grapheme to phoneme translation	Inner speech	Phonetic coding and articulation	Verbal working memory
Homophone judgement	Y	Y	Y	N	N
Rhyme judgement	Y	Y	Y	N	Y
Reading word aloud	Y	Y	N	Y	N
Sentence repetition	N	N	N	Y	Y

Y = yes; N = no.

all analyses. Structural analyses were done on the remaining 17 patients who completed both tasks. All areas listed in the lesion analyses below were in the left hemisphere.

Rhyme judgement

Performance on the rhyming task was significantly associated with lesions to an area extending from the left inferior frontal gyrus pars opercularis (BA 44) and pars triangularis (BA 45), posteriorly through the pre- and postcentral gyri into the anterior part of the supramarginal gyrus (BA 40) and the white matter medial to it ($t = 3.73$, $P < 0.05$ permutation correction). Highest Z-scores were found in the pars opercularis and supramarginal gyrus. All effects remained significant when adding time post stroke or age ($P < 0.05$, false discovery rate correction) with a strong trend ($P < 0.001$ uncorrected) in pars opercularis and supramarginal gyrus when lesion volume was added as a covariate.

When scores for reading aloud were added as a covariate to control for speech production ability, the association of impaired rhyming performance and damage to the inferior frontal gyrus (pars opercularis and pars triangularis, extending into the precentral gyrus), supramarginal gyrus (BA 40) and the white matter medial to supramarginal gyrus remained significant ($P < 0.05$, false discovery rate correction; Fig. 3). Likewise, when scores from the sentence repetition task were entered as a covariate to control for verbal working memory, the effect in inferior frontal gyrus (pars opercularis, extending posteriorly and medially into the white matter), as well as a small cluster of white matter superior and medial to the supramarginal gyrus remained significant ($P < 0.05$, false discovery rate correction). There was also a

strong trend for an association between performance and the supramarginal gyrus ($P < 0.001$, uncorrected).

Similar results were obtained when the left-handed patients were removed from the analysis. Specifically, lesions to the area extending from the left inferior frontal gyrus, through the pre- and postcentral gyri into the anterior part of the supramarginal gyrus and the white matter medial to it were significantly associated with performance ($t = 4.00$, $P < 0.05$, permutation correction), although the Z-scores obtained were lower this time.

Homophone judgement

The lesion sites associated with performance on the homophone task replicated those observed for the rhyming task. Specifically, significant effects were observed in the superior part of the inferior frontal gyrus pars opercularis and white matter medial to the supramarginal gyrus ($t = 3.85$, $P < 0.05$, permutation correction) and these effects remained significant after including time post stroke or age as covariates ($P < 0.05$, false discovery rate correction). The effect in the inferior frontal gyrus also remained significant after the left-handed patients were excluded ($t = 4.13$, $P < 0.05$, permutation correction) with a strong trend in the white matter medial to the supramarginal gyrus ($P < 0.001$ uncorrected). However, when reading aloud was added as a covariate, the significance of the effects were reduced below uncorrected level of $P < 0.001$: lesions to the pars opercularis were associated with performance on the homophone judgement task only at $P < 0.005$, uncorrected, while lesions to the white matter adjacent to the supramarginal gyrus were associated with performance only at $P < 0.01$, uncorrected.

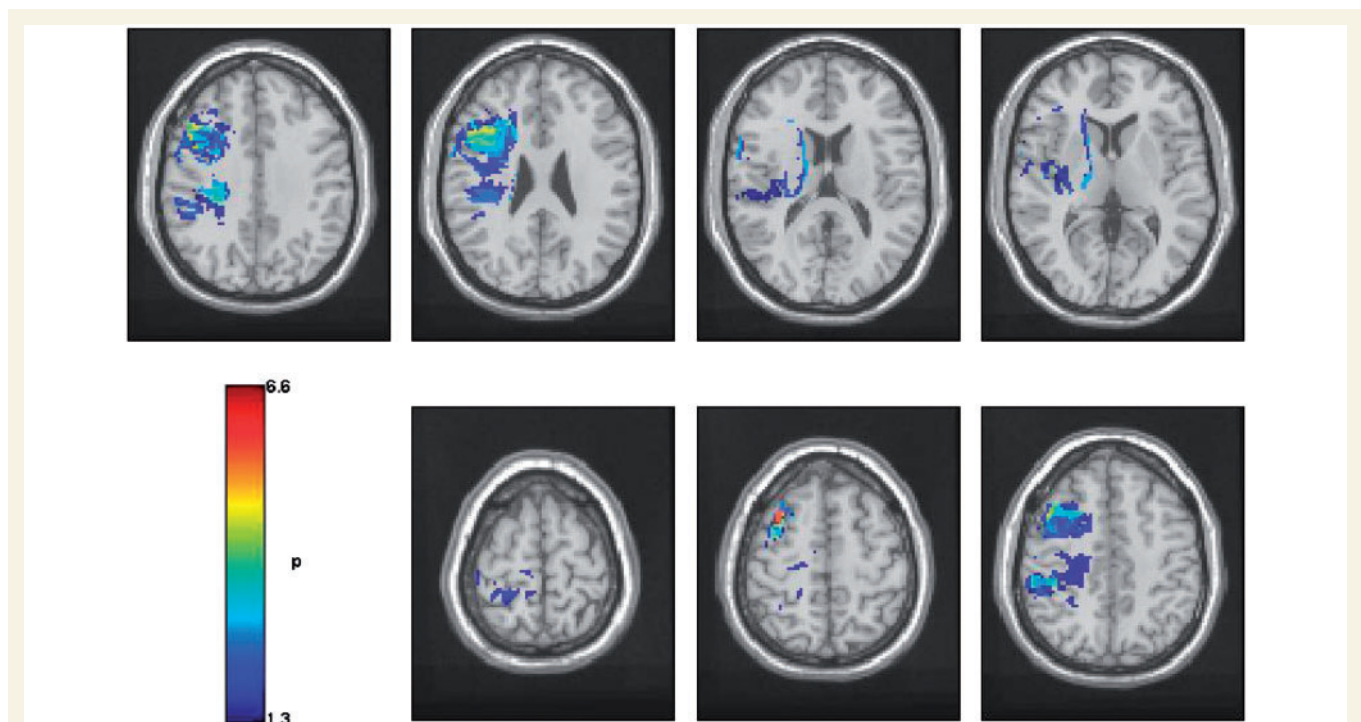


Figure 3 Voxel-based lesion symptom maps showing areas of significant association between lesion and performance on the rhyme judgement task after speech production has been controlled ($P < 0.05$, false discovery rate corrected). All P -values are in $-\log_{10}$.

Together these results suggest that difficulties with tasks requiring inner speech are associated with lesions to the inferior frontal gyrus pars opercularis (for both rhyme and homophone judgements) and with lesions to the supramarginal gyrus or the white matter adjacent to it (for rhyme judgement).

Discussion

This study is the first to look specifically at the relationship between inner speech, overt speech and lesion site in post-stroke aphasia, using structural analysis. It shows that lesions to the inferior frontal gyrus pars opercularis (BA 44) and supramarginal gyrus (BA 40) and its adjacent white matter are correlated with performance on a rhyming task, even when reading aloud or sentence repetition was factored out. Corresponding, albeit weaker, effects were observed for performance on homophone judgements. Together these findings support the main hypothesis, that structures in the left inferior frontal gyrus and inferior parietal lobe are related to inner speech. However, the left inferior frontal gyrus and inferior parietal lobe are anatomically large and functionally diverse regions. Our study adds information about the specific functions of these regions.

The result showing that the left inferior frontal gyrus is important for inner speech processing is consistent with functional imaging studies of inner speech, especially of rhyme judgement (Paulesu *et al.*, 1993; Pugh *et al.*, 1996; Lurito *et al.*, 2000; Poldrack *et al.*, 2001; Owen *et al.*, 2004; Hoeft *et al.*, 2007), and with other studies showing that the left inferior frontal gyrus, and particularly the pars opercularis, is involved in phonological processing (Mummery *et al.*, 1996; McDermott *et al.*, 2003; Burton *et al.*, 2005; Price *et al.*, 2005).

On the other hand, the functional imaging studies comparing inner and overt speech, which are most similar to our analysis, show more complex results. These studies used various types of tasks and the nature of these tasks might be a key to the interpretation of the results. The rhyme and homophone judgement tasks require a level of active 'use' of inner speech, in a way that one has to monitor, or listen to, one's own inner speech in order to successfully perform the task. This might be the case also in the semantic fluency task (Basho *et al.*, 2007) and when generating names of animals (Huang *et al.*, 2002), where the participant needs to keep track of the words already produced. In the case of word repetition (Shuster and Lemieux, 2005), letter naming (Huang *et al.*, 2002), silent reading (Bookheimer *et al.*, 1995) and counting (Ryding *et al.*, 1996), such monitoring of inner speech is less crucial for performance. Using terminology coined by Vigliocco and Hartsuiker (2002), it is suggested that the tasks used in this study, along with the semantic fluency task and generating animal names, most likely require 'conscious inner speech'. In contrast, the other tasks require only 'unconscious inner speech'. The neural correlates associated with these two types of inner speech can potentially be somewhat different. Re-examination of the results of Huang *et al.* (2002) shows that generating animal names, a task which requires a more conscious type of inner speech, produced high left inferior frontal gyrus activation, while naming letters, which requires a less conscious type

of inner speech, did not produce significant left inferior frontal gyrus activation. Therefore, it is suggested that the left inferior frontal gyrus is more closely related to conscious inner speech. Lastly, an interesting study evaluated neural correlates associated with verbal transformations (Sato *et al.*, 2004). 'Verbal transformation' refers to the phenomenon where a word is repeated rapidly, and after a while a new percept 'pops-out', and the participant starts perceiving a word that is different from the one perceived initially. For example, if the word 'life' is repeated rapidly, it might after some time sound like 'fly'. Sato *et al.* (2004) compared two conditions: in the first condition participants were asked to simply repeat the word, while in the second they were asked to pay attention to the moment in which a verbal transformation occurs. In this way, the authors created two conditions: more unconscious inner speech (the former) versus more conscious inner speech (the latter), using the same stimuli. Comparing the two conditions directly (attending to the verbal transformation > repetition), they found that conscious inner speech was significantly correlated with activation in the left inferior frontal gyrus and left supramarginal gyrus, as well as other regions (anterior part of the right cingulate gyrus, bilateral cerebellum and left superior temporal gyrus). It should be clarified that it is not proposed that a clear-cut dissociation between conscious and unconscious inner speech can be made. Rather, it is suggested that different tasks require different levels of inner speech, and that this, in turn, influences the involvement of the left inferior frontal gyrus.

Our study also identified the white matter adjacent to the supramarginal gyrus, as important for both our inner speech tasks. These white matter areas are likely to be part of the dorsal route for language. Clearly, a technique which specializes in imaging white matter (e.g. diffusion tensor imaging) could shed more light on the exact anatomy and function of these white matter tracts (Geva *et al.*, 2011). However, our results suggest that the dorsal route for language is required for inner speech. The dorsal language route is mainly composed of the arcuate fasciculus and the superior longitudinal fasciculus [see Geva *et al.* (2011) for a review of recent anatomical studies of the dorsal language route]. Previous studies of the dorsal language route in patients with aphasia have shown that the arcuate fasciculus (Breier *et al.*, 2008; Fridriksson *et al.*, 2010) or the superior longitudinal fasciculus (Breier *et al.*, 2008) are crucial for repetition. However, case studies show that lesions to the arcuate fasciculus do not always result in conduction aphasia (Mori and van Zijl, 2002; Selnes *et al.*, 2002; Bernal and Ardila, 2009) and conversely, that behavioural phenotypes of conduction aphasia can arise from lesions in regions outside the arcuate fasciculus (Benson and Ardila, 1996). Wise *et al.* (2001) have suggested that the junction between the posterior supratemporal region and the inferior parietal lobe acts as a centre for binding speech perception and speech production, or lexical recall. This is relevant for tasks involving repetition and auditory comprehension, and also for inner speech production. And indeed, the involvement of the supramarginal gyrus in repetition has been demonstrated previously (Anderson *et al.*, 1999; Quigg *et al.*, 2006). Diffusion tensor imaging studies demonstrated that the supramarginal gyrus and BA 44 are connected via a direct connection (Catani *et al.*, 2005; Frey *et al.*, 2008), as

well as via indirect connections through the posterior superior temporal gyrus (Parker *et al.*, 2005; Friederici, 2009). Recently, it was shown that the pars opercularis shows functional connectivity to the supramarginal gyrus during resting state, and it was suggested that the two regions function together as a phonological processing system (Xiang *et al.*, 2010). The role of the dorsal route in supporting inner speech might be to transfer the output phonological code from anterior areas such as the left inferior frontal gyrus to posterior regions, where it is further processed. It should be noted that previous studies emphasized a specific functional directionality of the dorsal route, in which processing advances from posterior to anterior regions, supporting repetition, language acquisition and monitoring of overt speech (Catani *et al.*, 2005; Hickok and Poeppel, 2007; Saur *et al.*, 2008; Friederici, 2009; Agosta *et al.*, 2010). A study of the macaque brain showed that Area 44, which is homologous to BA 44 in the human brain, receives input from area PFG of the inferior parietal lobule, which is homologous to the caudal part of the supramarginal gyrus in humans (Petrides and Pandya, 2009). Since this study or others did not examine whether the pars opercularis sends fibres to the supramarginal gyrus, in the monkey's brain, the possibility that such connections exist cannot be ruled out, and indeed, most cortico-cortical connections in mammals are reciprocal. Matsumoto *et al.* (2004) used electrodes to directly stimulate anterior and posterior cortical regions and record evoked potentials in humans. Language areas were defined as those which, when stimulated, impair sentence reading in the individual patient. Anterior regions included Broca's area or adjacent regions and posterior regions included the supramarginal gyrus, the middle and posterior superior temporal gyrus, and the adjacent middle temporal gyrus. As expected, stimulation of posterior language areas resulted in evoked potentials in anterior language areas, supporting the idea of processing progressing from posterior to anterior. However, stimulation of anterior regions also resulted in evoked potentials in all posterior regions tested, including the supramarginal gyrus, middle and posterior superior temporal gyri and the adjacent middle temporal gyrus. In summary, although studies emphasize a propagation of information in the dorsal route from posterior to anterior parts, it is possible that some reciprocal fibres in this pathway send information in the other direction, and these might be essential for inner speech production.

In our study, the inner and overt tasks involved reading of irregular words. The use of irregular words was crucial to the study, since it forced the readers to use inner speech, rather than making the decision based on the orthographic form of the word. Patients with poor performance on both tasks might therefore have deficits in accessing the phonological form of irregular words from written text, as is the case in surface alexia (Benson and Ardila, 1996). However, surface alexia cannot explain effects that were greater for the inner speech tasks than the overt speech tasks.

It should be noted that the source of the inner speech impairment in each patient was not characterized here. In a previous paper (Geva *et al.*, 2011), we suggested that inner speech deficits can arise from impairment to the production system, the comprehension system or the connection between the two. The anatomical results presented here suggest that the group of patients tested is a mixed one and likely to present all sources of

impairments to inner speech. However, in our cohort there were no patients with severe global aphasia and therefore the results cannot be necessarily generalized to those patients.

This study's main potential caveat is the inclusion of participants who were not strongly right-handed. However, all patients exhibited language impairments following a left middle cerebral artery stroke. Moreover, an analysis excluding those patients showed similar results. A second caveat is the small number of patients. Although previous papers have suggested which statistics should be employed with small numbers of patients (Medina *et al.*, 2010), thus enabling the interpretation of results drawn from small datasets, a replication of these results with a larger dataset is needed. Lastly, we identified the main anatomical correlates of inner speech using permutation testing, a recommended method of statistical correction for multiple comparisons (Medina *et al.*, 2010). Then once regions were identified we reported *post hoc* tests of the effects using false discovery rate correction. While many advocate the use of permutation testing when one variable is analysed, there is a range of possible methods when a model includes multiple variables, as we used here. For example, Nichols *et al.* (2008) have suggested a method that entails first regressing out the effect of the covariate of no interest, and then running the permutation test on the residuals, using the covariate of interest. Future work might help define which statistical procedure is most appropriate for different datasets.

The results of this study may influence the construction of future language imaging paradigms. Huang *et al.* (2002) stated that 'it is incorrect to view the neural substrates of silent and overt speech as the same up to the execution of motor movements, and it is therefore inappropriate to use silent speech as a motion-free substitute for overt speech in studies of language production' (Huang *et al.*, 2002, p. 50), and our results support this statement. It is, therefore, suggested that inner and overt speech tasks have differences as well as overlaps in their brain localization. Hence, using inner speech tasks to study overt speech production may well result in misleading conclusions. Secondly, our results might have implications for the use of imaging paradigms as preoperative evaluations for tumour and epileptic patients due to undergo brain surgery. Functional MRI paradigms often employ inner speech tasks, while preoperative assessments (using the invasive Wada test or direct cortical stimulation) usually employ overt speech tasks (Foki *et al.*, 2008). In accordance with Foki *et al.* (2008), we suggest that replacing the Wada test and direct electrostimulation, with functional MRI, requires the understanding of the anatomical differences between inner and overt speech. Lastly, together with previous behavioural results (Geva *et al.*, 2011), this study suggest that inner and overt speech cannot be treated as one and the same. This might influence diagnosis and prognosis procedures of patients with post-stroke aphasia.

Conclusion

This study investigated the neural correlates of inner speech using a structural analysis method. The left inferior frontal gyrus (pars opercularis), left supramarginal gyrus and white matter regions adjacent to the supramarginal gyrus were found to be involved

in inner speech processing. These regions are part of the dorsal route for language. It is, therefore, suggested that inner speech is produced by frontal regions (BA 44) and it is then transferred via the arcuate fasciculus to posterior regions that link speech production to speech comprehension. By showing that inner speech cannot be described as simply overt speech without a motor component, this study has implications to the construction of future language imaging studies and clinical practice.

Acknowledgements

We thank Nicola Lambert, Helen Walker, Sinead Stringwell and Helen Palmer for helping with recruitment of patients, and Tulasi Marrapu for helping with the analysis.

Funding

Medical Research Council (grant number G0500874 to J.-C.B. and E.A.W.); Pinsent-Darwin Fellowship, Wingate scholarship, The Cambridge Overseas Trust and B'nai Brith Scholarship (to S.G.); Cambridge Comprehensive Biomedical Research Centre (BRC) (P.S.J. and E.A.W.); UK National Institute of Health Research (NIHR) (E.A.W.).

References

- Agosta F, Henry RG, Migliaccio R, Neuhaus J, Miller BL, Dronkers NF, et al. Language networks in semantic dementia. *Brain* 2010; 133: 286–99.
- Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, et al. Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain Lang* 1999; 70: 1–12.
- Baddeley A, Hitch G. Working memory. In: Bower GH, editor. *The psychology of learning and motivation: advances in research and theory*. New York: Academic Press; 1974. p. 47–89.
- Ballet G. *Le langage interieur et les diverses formes de l'aphasie*. Paris: Ancienne Librairie Germer Bailliere; 1886.
- Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. *Neuroimage* 1999; 10: 642–57.
- Basho S, Palmer ED, Rubio MA, Wulfeck B, Muller RA. Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. *Neuropsychologia* 2007; 45: 1697–706.
- Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, et al. Voxel-based lesion-symptom mapping. *Nat Neurosci* 2003; 6: 448–50.
- Benson DF, Ardila A. *Aphasia: a clinical perspective*. Oxford: Oxford University Press; 1996.
- Bernal B, Ardila A. The role of the arcuate fasciculus in conduction aphasia. *Brain* 2009; 132: 2309–16.
- Besner D, Davies J, Daniels S. Reading for meaning: the effects of concurrent articulation. *Quart J Exp Psychol* 1981; 33: 415–37.
- Blonskii PP. *Memory and thought. Selected works in psychology*. Moscow: Prosveshchenie Press; 1964.
- Bookheimer SY, Zeffiro TA, Blaxton T, Gaillard W, Theodore W. Regional cerebral blood flow during object naming and word reading. *Human Brain Mapp* 1995; 3: 93–106.
- Breier JI, Hasan KM, Zhang W, Men D, Papanicolaou AC. Language dysfunction after stroke and damage to white matter tracts evaluated using diffusion tensor imaging. *Am J Neuroradiol* 2008; 29: 483–7.
- Brown GDA. Phonological coding in rhyming and homophony judgement. *Acta Psychol* 1987; 65: 247–62.
- Burgess P, Shallice T. *The Hayling and Brixton Tests*. Test manual. Bury St Edmunds, UK: Thames Valley Test Company; 1997.
- Burton MW, LoCasto PC, Krebs-Noble D, Gullapalli RP. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage* 2005; 26: 647–61.
- Catani M, Jones DK, Ffytche DH. Perisylvian language networks of the human brain. *Ann Neurol* 2005; 57: 8–16.
- Corcoran DWJ. An acoustic factor in letter cancellation. *Nature* 1966; 210: 658.
- Dabul BL. *Apraxia Battery for Adults (ABA)*. USA: C.C. Publications, Inc; 1979.
- Egger V. *La Parole Interieure*. Paris: Ancienne Librairie Germer Bailliere; 1881.
- Feinberg TE, Gonzalez Rothi LJ, Heilman KM. 'Inner speech' in conduction aphasia. *Arch Neurol* 1986; 43: 591–3.
- Foki T, Gartus A, Geissler A, Beisteiner R. Probing overtly spoken language at sentential level - A comprehensive high-field BOLD-fMRI protocol reflecting everyday language demands. *Neuroimage* 2008; 39: 1613–24.
- Frey S, Campbell JSW, Pike GB, Petrides M. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci* 2008; 28: 11435–44.
- Fridriksson J, Kjartansson O, Morgan PS, Hjaltason H, Magnusdottir S, Bonilha L, et al. Impaired speech repetition and left parietal lobe damage. *J Neurosci* 2010; 30: 11057–61.
- Friederici AD. Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci* 2009; 13: 175–81.
- Frings M, Dimitrova A, Schorn CF, Elles HG, Hein-Kropp C, Gizewski ER, et al. Cerebellar involvement in verb generation: An fMRI study. *Neurosci Lett* 2006; 409: 19–23.
- Geva S, Bennett S, Warburton EA, Patterson K. Discrepancy between inner and overt speech: Implications for post stroke aphasia and normal language processing. *Aphasiology* 2011; 25: 323–43.
- Geva S, Correia M, Warburton EA. Diffusion tensor imaging in the study of language and aphasia. *Aphasiology* 2011; 25: 543–58.
- Gracco VL, Tremblay P, Pike B. Imaging speech production using fMRI. *Neuroimage* 2005; 26: 294–301.
- Hickok G, Poeppel D. Opinion - The cortical organization of speech processing. *Nat Rev Neurosci* 2007; 8: 393–402.
- Hoeft F, Meyler A, Hernandez A, Juel C, Taylor-Hill H, Martindale JL, et al. Functional and morphometric brain dissociation between dyslexia and reading ability. *Proc Natl Acad Sci USA* 2007; 104: 4234–9.
- Howard D, Franklin S. Memory without rehearsal. In: Vallar G, Shallice T, editors. *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press; 1990. p. 287–318.
- Huang J, Carr TH, Cao Y. Comparing cortical activations for silent and overt speech using event-related fMRI. *Hum Brain Mapp* 2002; 15: 39–53.
- Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. *Cognition* 2004; 92: 101–44.
- Johnston RS, McDermott EA. Suppression effects in rhyme judgement tasks. *Quart J Exp Psychol Hum Exp Psychol* 1986; 38: 111–24.
- Kay J, Coltheart M, Lesser R. *Psycholinguistic assessment of language processing*. Hove, UK: Psychology Press; 1992.
- Kimberg DY, Coslett HB, Schwartz MF. Power in voxel-based lesion-symptom mapping. *J Cogn Neurosci* 2007; 19: 1067–80.
- Kleiman GM. Speech recoding in reading. *J Verbal Learn Behav* 1975; 14: 323–39.
- Lurito JT, Kareken DA, Lowe MJ, Chen SHA, Mathews VP. Comparison of rhyming and word generation with fMRI. *Hum Brain Mapp* 2000; 10: 99–106.
- Marshall RC, Neuburger SI, Phillips DS. Verbal self-correction and improvement in treated aphasic clients. *Aphasiology* 1994; 8: 535–47.
- Marshall RC, Rappaport BZ, Garciabunuel L. Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain Lang* 1985; 24: 297–313.

- Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and fronto-temporal dementia. *Neurology* 2000; 55: 1613–20.
- Matsumoto R, Nair DR, LaPresto E, Najm I, Bingaman W, Shibasaki H, et al. Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain* 2004; 127: 2316–30.
- McDermott KB, Petersen SE, Watson JM, Ojemann JG. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 2003; 41: 293–303.
- Medina J, Kimberg DY, Chatterjee A, Coslett HB. Inappropriate usage of the Brunner-Munzel test in recent voxel-based lesion-symptom mapping studies. *Neuropsychologia* 2010; 48: 341–3.
- Meyers JE, Meyers KR. *Rey Complex Figure Test and Recognition Trial [Professional Manual]*. Odessa, Florida, USA: Psychological Assessment Resource, Inc; 1995.
- Mori S, van Zijl PCM. Fiber tracking: principles and strategies - A technical review. *NMR Biomed* 2002; 15: 468–80.
- Morin A, Michaud J. Self-awareness and the left inferior frontal gyrus: inner speech use during self-related processing. *Brain Res Bull* 2007; 74: 387–96.
- Mummery CJ, Patterson K, Hodges JR, Wise RJS. Generating 'tiger' as an animal name or a word beginning with T: differences in brain activation. *Proc R Soc Lond Ser B Biol Sci* 1996; 263: 989–95.
- Nichols T, Ridgway G, Webster M, Smith S. GLM permutation - non-parametric inference for arbitrary general linear models. In Abstract proceedings of 14th Annual Meeting of the Organization for Human Brain Mapping. Australia: Melbourne; 2008.
- Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 2008; 106: 528–37.
- Oppenheim GM, Dell GS. Motor movement matters: the flexible abstractness of inner speech. *Memory Cogn* 2010; 38: 1147–60.
- Owen WJ, Borowsky R, Sarty GE. fMRI of two measures of phonological processing in visual word recognition: ecological validity matters. *Brain Lang* 2004; 90: 40–6.
- Ozdemir R, Roelofs A, Levelt WJM. Perceptual uniqueness point effects in monitoring internal speech. *Cognition* 2007; 105: 457–65.
- Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley WM, Petersen SE. An event-related fMRI study of overt and covert word stem completion. *Neuroimage* 2001; 14: 182–93.
- Parker GJM, Luzzi S, Alexander DC, Wheeler-Kingshott CAM, Clecarelli O, Ralph MAL. Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage* 2005; 24: 656–66.
- Paulesu E, Frith CD, Frackowiak RSJ. The neural correlates of the verbal component of working memory. *Nature* 1993; 362: 342–5.
- Peck KK, Moore AB, Crosson BA, Gaiefsky M, Gopinath KS, White K, et al. Functional magnetic resonance imaging before and after aphasia therapy - Shifts in hemodynamic time to peak during an overt language task. *Stroke* 2004; 35: 554–9.
- Petrides M, Pandya DN. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *Plos Biol* 2009; 7: 16.
- Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, et al. Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *J Cogn Neurosci* 2001; 13: 687–97.
- Postma A, Noordanus C. Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Lang Speech* 1996; 39: 375–92.
- Price CJ, Devlin JT, Moore CJ, Morton C, Laird AR. Meta-analyses of object naming: Effect of baseline. *Hum Brain Mapp* 2005; 25: 70–82.
- Pugh KR, Shaywitz BA, Shaywitz SE, Constable RT, Skudlarski P, Fulbright RK, et al. Cerebral organization of component processes in reading. *Brain* 1996; 119: 1221–38.
- Quigg M, Geldmacher DS, Elias WJ. Conduction aphasia as a function of the dominant posterior perisylvian cortex - Report of two cases. *J Neurosurg* 2006; 104: 845–8.
- Richardson JTE. Phonology and reading: the effects of articulatory suppression upon homophony and rhyme judgements. *Lang Cogn Proces* 1987; 2: 229–44.
- Rorden C, Karnath HO, Bonilha L. Improving lesion-symptom mapping. *J Cogn Neurosci* 2007; 19: 1081–8.
- Rudrauf D, Mehta S, Bruss J, Tranel D, Damasio H, Grabowski TJ. Thresholding lesion overlap difference maps: application to category-related naming and recognition deficits. *Neuroimage* 2008; 41: 970–84.
- Ryding E, Bradvik B, Ingvar DH. Silent speech activates prefrontal cortical regions asymmetrically, as well as speech-related areas in the dominant hemisphere. *Brain Lang* 1996; 52: 435–51.
- Sato M, Baciú M, Loevenbruck H, Schwartz JL, Cathiard MA, Segebarth C, et al. Multistable representation of speech forms: A functional MRI study of verbal transformations. *Neuroimage* 2004; 23: 1143–51.
- Saur D, Kreher BW, Schnell S, Kummerer D, Kellmeyer P, Vry MS, et al. Ventral and dorsal pathways for language. *Proc Natl Acad Sci USA* 2008; 105: 18035–40.
- Selnes OA, van Zijl PCM, Barker PB, Hillis AE, Mori S. MR diffusion tensor imaging documented arcuate fasciculus lesion in a patient with normal repetition performance. *Aphasiology* 2002; 16: 897–901.
- Shuster LI, Lemieux SK. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain Lang* 2005; 93: 20–31.
- Sokolov AN. *Inner speech and thought* (Onischenko GT, Trans.). New York, London: Plenum Press; 1972.
- Swinburn K, Porter G, Howard D. *Comprehensive aphasia test*. East Sussex, UK: Psychology Press; 2004.
- Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. *Psychol Bull* 2002; 128: 442–72.
- Vygotsky LS. *Thought and Language* (Hanfmann E and Vakar G, Trans.). Cambridge, Massachusetts M.I.T. Press; 1962.
- Wilding JM, White W. Impairment of rhyme judgments by silent and overt articulatory suppression. *Quart J Exp Psychol Hum Exp Psychol* 1985; 37: 95–107.
- Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural subsystems within 'Wernicke's area'. *Brain* 2001; 124: 83–95.
- Xiang HD, Fonteijn HM, Norris DG, Hagoort P. Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral Cortex* 2010; 20: 549–60.