Reading fluent speech from talking faces:

Typical brain networks and individual differences

Deborah A Hall, Clayton Fussell, A Quentin Summerfield

MRC Institute of Hearing Research,
University Park,
Nottingham,
NG7 2RD, UK
Tel : +44 (115) 922 3431
Fax : +44 (115) 951 8503
Email : d.hall@ihr.mrc.ac.uk

Abstract

Listeners are able to extract important linguistic information by viewing the talker's face – a process known as 'speechreading'. Previous studies of speechreading present small closed sets of simple words and their results indicate that visual speech processing engages a wide network of brain regions in the temporal, frontal, and parietal lobes that are likely to underlie multiple stages of the receptive language system. The present study further explored this network in a large group of subjects by presenting naturally-spoken sentences which tap the richer complexities of visual speech processing. Four different baselines (blank screen, static face, non-linguistic facial gurning, and auditory speech) enabled us to determine the hierarchy of neural processing involved in speechreading and to test the claim that visual input reliably accesses sound-based representations in the auditory cortex.

In contrast to passively viewing a blank screen, the static-face condition evoked activation bilaterally across the border of the fusiform gyrus and cerebellum, and in the medial superior frontal gyrus and left precentral gyrus ($P < 0.05$, whole brain corrected). With the static face as baseline, the gurning face evoked bilateral activation in the motion-sensitive region of the occipital cortex, whereas visual speech additionally engaged middle temporal

gyrus, inferior and middle frontal gyri, and the inferior parietal lobe, particularly in the left hemisphere. These latter regions are implicated in lexical stages of spoken language processing. While auditory speech generated extensive bilateral activation across both superior and middle temporal gyri, the group-averaged pattern of speechreading activation failed to include any auditory regions along the superior temporal gyrus, suggesting that fluent visual speech does not always involve sound-based coding of the visual input.

An important finding from the individual subject analyses was that activation in the superior temporal gyrus did reach significance ($P < 0.001$, small-volume corrected) for a subset of the group. Moreover, the extent of the left-sided superior temporal gyrus activity was strongly correlated with speechreading performance. Skilled speechreading was also associated with activations and deactivations in other brain regions suggesting that individual differences reflect the efficiency of a circuit linking sensory, perceptual, memory, cognitive, and linguistic processes rather than the operation of a single component process.

Introduction

Speechreading is the ability to understand a talker by viewing the movements of their lips, teeth, tongue, and jaw. Visual speech therefore conveys specific linguistic information that is separate from the analysis of facial features and of non-meaningful movements of the face (Figure 1). In isolation, speechreading is rarely perfect because some phonetic distinctions between consonants do not have visible articulatory correlates and fluent speech contains coarticulation between words. Nevertheless linguistic facial movements exert an automatic influence on the perception of heard speech; famously demonstrated by the fusion illusion in which, for example, the congruent presentation of a heard syllable [ba] and a seen utterance [ga] generates the subjective impression of hearing [da] (McGurk & MacDonald, 1976). In normal listening situations, visual information facilitates perceptual accuracy by providing supplementary information about the spatial location of the talker and about the

segmental elements of the speech acoustic waveform, as well as complementary information about consonant place of articulation that is most susceptible to acoustic distortion (Summerfield, 1987). Thus, linguistic facial movements are particularly informative for speech perception when the auditory signal is degraded by noise (Sumby & Pollack, 1954), reverberation, or the distortions introduced by a sensori-neural hearing impairment (Jeffers & Barley, 1971). Macleod and Summerfield (1987; 1990) demonstrated that the comprehension benefit at poor acoustic signal-to-noise ratios from seeing the talker's face was closely related to the success with which those subjects were able to perform a comparable silent speechreading test ($r=0.9$, $P < 0.01$). Therefore, silent speechreading and audiovisual speech perception are likely to share common processes of visual analysis. As a consequence, tests of speechreading have often been used to probe the patterns of brain activation that support audiovisual speech processing. The results of these neuroimaging studies are reviewed in the following section.

Neural underpinnings of speechreading simple word lists

The neural basis of speechreading has been measured by presenting subjects with simple word lists such as the numbers one to ten or other closed sets of high-frequency words. One of the first studies to elucidate some of the brain areas specifically involved in speechreading presented five normally-hearing individuals with auditory speech, visual speech, pseudo-speech, closed-mouth gurning, static face, and silent baseline conditions (Calvert et al., 1997). Relative to the static face, visual speech engaged extra-striate visual cortex implicated in the detection of visual motion and the superior temporal gyrus traditionally viewed as unimodal auditory cortex. Parts of the auditory cortex were claimed to be involved in encoding at a pre-lexical, phonological stage of visual speech processing because they were also engaged by pseudo-speech, but not by the gurning face. At the time of publication, fMRI methodology was at an early stage of development: the imaging view was

limited to a portion of the brain, the activation maps were those of a simple composite group average and the functional localisation was approximate. Nevertheless, the finding that visual speech elicits superior temporal gyrus activation has been replicated many times since (Bernstein et al., 2002; Calvert & Campbell, 2003; Calvert et al., 2000; Campbell et al., 2001, MacSweeney et al., 2002), and even in the absence of scanner acoustic noise (MacSweeney et al., 2000; Paulesu et al., 2003). Whole brain scanning has revealed that visual speech generates robust activation in the inferior frontal gyrus (BA 44 and 45), occipito-temporal junction (BA 37), middle temporal gyrus (BA 21), and superior temporal sulcus, but generally with greater activation on the left than the right side. These regions are implicated in traditional speech-processing circuitry and can be engaged even when syllables, not words or sentences, are presented (Calvert & Campbell, 2003). Speechreading activation typically also includes the precentral gyrus (BA 6) and parietal lobe (BA 7 and 40) in both hemispheres. It is proposed that the processes underlying perception and action might share a common representational framework such that premotor, parietal, and inferior frontal brain regions could support the perception of meaningful actions (Decety & Grezes, 1999). Indeed, observing speech-related lip movements has been shown to enhance the excitability of the primary motor units underlying speech production, particularly those in the left hemisphere (Watkins et al., 2003).

Individual differences in speechreading ability

Closed sets of high-frequency monosyllabic or bisyllabic words are relatively easy to speechread and so previous neuroimaging studies tend to report performance levels that are close to ceiling. However, for fluent speech, large individual differences in speechreading ability exist in both normally-hearing and hearing-impaired populations. Fluent visual speech contains many ambiguities because some phonetic distinctions do not have visible articulatory correlates and there is coarticulation across word boundaries. Speechreading ability can be measured by scoring the percentage of content words correctly reported in short sentences

(Macleod & Summerfield, 1990). Scores on such tests often range from less than 10% correct to over 70% correct among any group screened to have normal vision and similar hearing status and age. For 20 normally-hearing adults, Macleod and Summerfield (1987) reported a spread in performance across the bottom 50% range of scores on a subset of the British-English sentences developed by Bamford et al. (1979). For 68 deaf children aged 9 to 19 years, Heider and Heider (1940) reported a spread of scores from 11 to 93% on a story-like sentence comprehension test.

Since the inter-subject variability in speechreading is much greater than in auditory speech perception (Macleod and Summerfield, 1987), individual differences must reflect a process other than normal variation in speech-perception abilities. Speechreading fundamentally engages sensory, perceptual, memory, cognitive, and linguistic processes. Individual ability might reasonably depend on the operation of these processes collectively. However, identification of the core subskills that underlie individual differences in speechreading ability has so far remained elusive. There is evidence that the speed of neural transmission in the visual cortex might be implicated. A series of studies have shown that the latency of certain evoked responses to a brief light flash measured from the scalp was negatively correlated with speechreading ability. Initial studies identified an important bilateral visual evoked deflection at 130 ms after stimulus onset (Shepherd et al, 1977; Shepherd, 1982), while a subsequent study identified an additional earlier deflection at 16 ms (Samar & Sims, 1983). These findings suggest a role for general sensory function or regulation of visual attention in mediating individual differences in speechreading. However, more recently it has been shown that the correlation effect is influenced by the hearing status of the subject and by the stimulus predictability (Rönneburg et al., 1989; Samar & Sims, 1984) and may be present only for the amplitude of the negative deflection not its latency (Rönneburg et al., 1989). Thus, if visual processing speed is related to speechreading ability it is not a simple or straightforward index.

Speechreading ability was initially thought to implicate general intelligence and verbal reasoning, but these cognitive skills correlate poorly with speechreading performance (Jeffers & Barley, 1971; Macleod & Summerfield, 1990). Furthermore, speechreading ability does not merely reflect the ability to parse a stream of articulatory gestures because performance is also correlated with the ability to speechread isolated words and syllables (Summerfield, 1991). Better speechreaders are likely to more effectively employ visual analytic skills, linguistic context, and other strategies for rejecting less likely interpretations of ambiguous words, yet effective speechreading strategies are notoriously difficult to teach and learn (Binnie, 1977; Heider & Heider, 1940) and so the behavioural basis for individual differences is still uncertain. Summerfield (1991) noted that, despite being highly motivated, even the best totally deaf and hearing-impaired subjects often perform only as well as the best subjects with normal hearing on laboratory tests of fluent speechreading; indicating that speechreading is an independent trait. Summerfield concludes that "good speechreaders are born, not made" (p122).

To date there have been no systematic studies of the neural basis for individual differences in speechreading ability. A preliminary study found some evidence that poor speechreaders displayed less activation in middle temporal gyrus, but with small numbers of subjects this was not a very reliable effect (Ludman et al., 2000). More recently, Paulesu et al. (2003) exploited subjects' improvement in performance with successive repetitions of the same 24 words to give a spread of speechreading scores against which he could regress rCBF in a within- and between- subjects design. Negative correlations between activation and performance were found in many brain areas including inferior frontal gyrus, inferior parietal lobe, temporal lobe, and the temporo-occipital junction. However, the degree to which these activations reflect either training or true individual differences in ability is unclear because the design does not distinguish activation related to learning the set of words from activation related to individual differences in speechreading.

The neural basis of speechreading has so far been measured by presenting subjects with closed sets of simple words such as the numbers one to ten or other high-frequency words. Such stimuli do not tap all of the mental operations that are required for the perception of fluent speech where a continuous articulatory stream must be parsed into words, aided by prosodic cues and syntactic and semantic context. There is supporting evidence from auditory speech comprehension that the temporo-frontal network is recruited differently for spoken word lists than for fluent speech, possibly due to syntactic and prosodic differences (Friederici et al., 2000; Hickok & Poeppel, 2000).

In the current experiment, we obtained behavioural and physiological measures of speechreading in a large group of participants. Our aims were threefold: i) to reliably distinguish those brain regions involved in the sensory and higher-order aspects of speechreading sentences, ii) to test whether visual speech reliably accesses the same cortical regions as does auditory speech, especially in the region of the auditory cortex and iii) to identify the neural markers for speechreading proficiency. The aims were achieved by designing a visual and an auditory speech task in which participants were required to parse short spoken sentences (IHR number sentences) and identify the number embedded within each one.

Results

Speechreading ability

For the 33 participants, scores for the proportion of keywords correctly identified on the speechreading screening test ranged from 7% to 89% (mean = 39%, s.d. = 20%). When measured during the fMRI experiment, speechreading performance on the IHR number sentences spanned a similarly broad range (22-89%) with a mean of 51% and standard deviation of 14% (Figure 2). For the 33 participants, a strong correlation existed between the two measures of speechreading ability (Pearson $r = 0.80$, $P<0.01$). Therefore, we are confident

that the IHR number sentence task provided a sufficiently sensitive measure of speechreading ability and required subjects to comprehend the sentence rather than simply picking out the number word. Scores acquired during the fMRI study revealed that the auditory speech task was much easier than the visual speech task ($t [32] = 23.0$, $P < 0.001$) and subjects generally performed at ceiling (mean = 97%, s.d. = 7%).

Basic networks of activation

To reveal the brain regions engaged by speechreading and to make inferences about their functional role, we make the assumption of pure insertion. Our experiment contained tasks with varying levels of cognitive complexity. By contrasting these tasks in a pairwise manner, we can functionally decompose speechreading into the following components: i) general face analysis (static face - blank screen), ii) dynamic face analysis (gurning - static face), and iii) phonological, lexical, syntactic, and semantic (i.e. linguistic) analysis (visual speech - gurning). By the same logic of subtraction, we can similarly decompose the anatomical basis of these three speechreading components. In addition, by investigating the pattern of conjoint activation by visual and auditory speech, we can infer what mental processes might be commonly engaged by both tasks regardless of their different input modalities. Since the speech stimulus is constant across modalities, but the difficulty and success of comprehension is not, the experimental design offers a more valid comparison between regions of shared sensory or perceptual activation rather than shared cognitive or linguistic activation. The activation differences generated by these cognitive subtractions are reported in Tables 1 and 2. Our approach uses stringent statistical criteria for determining levels of significance so that we can draw conclusions about *typical* patterns of activation.

Static face – blank screen

This contrast activated brain areas that are represented in green in Figure 3A. Most notably, bilateral patterns of activation were found across the border between the fusiform gyrus (BA 18/19) and the cerebellum. The mid-fusiform gyrus on the ventral surface of the occipito-temporal cortex contains a face-selective region which plays a role in the perceptual processing of face stimuli (Grill-Spector et al., 2004). The band of bilateral activation observed in our data probably overlaps with the fusiform face area but, given the spatial normalisation inaccuracy in this region, it is difficult to conclude with certainty. Activation was also revealed in the medial superior frontal gyrus (BA 6) and precentral gyrus (BA 4). These areas are implicated in the motor response phase of the face task namely in the preparation, control, and production of signal-triggered finger presses, especially since the task required a sequence of movements (Tanji et al., 1996). The medial superior frontal gyrus and precentral gyrus were also activated by other pairwise contrasts where the motor requirements of the two conditions were not equally matched.

Gurning – static face

Compared with the static face, the gurning face elicited bilateral activation in extrastriate cortex (BA 19) extending to the temporo-occipital junction (Figure 3A, in blue). Activated voxels within this region fall in the vicinity of other group-averaged neuroimaging data that functionally define this region as V5(MT) – the motion-sensitive region (Hasnain et al., 1998). Specifically, the co-ordinates -48 -72 -4 mm and 44 -72 -12 mm are located 9 mm distant from those V5 centroids reported by Hasnain et al. (1998). This result is in line with previous findings that any sort of face motion processing (non-linguistic or otherwise) involves generic visual motion systems (Calvert et al., 1997).

Visual speech – gurning

In this contrast, meaningful facial gestures engaged additional regions of activation not seen in the previous contrasts (Figure 3A in red). These regions include inferior (BA 44/45/47) and middle (BA 9) frontal gyri, middle temporal gyri (BA 21), inferior parietal lobe (BA 7), and thalamus. The middle temporal gyrus activation extended up to the superior temporal sulcus in both hemispheres, but did not spread above this.

A striking observation was the left-sided emphasis in the pattern of activation; more than twice as many voxels in the left than in the right hemisphere reached significance. Such asymmetry is indicative of a language-based process. The network of activation observed here overlaps with those neural circuits involved in spoken language comprehension, suggesting many of these brain areas play a role in generic linguistic processes. An extensive left-sided distribution of speechreading activation across the inferior and middle frontal gyri has also been observed using a comparison between speechread real words and backwards words (Paulesu et al., 2003), but was not revealed by Calvert's comparison between talking and gurning lips (Calvert et al., 1997).

Auditory speech – silence

When contrasted against the silent resting baseline, auditory speech generated widespread activation in the bilateral temporal cortex (Figure 3B in green). Unlike the visual speech contrast, processing auditory speech was supported by the superior temporal gyrus (BA 41, 42, 22) as well as the middle temporal gyrus (BA 21). Compared with watching a talker's face, listening to the spoken sentences elicited much less frontal activation indicating a certain automaticity of the linguistic process, especially since the task did not accentuate any strategic or memory load (Friederici, 2002).

A common neural basis for processing visual and auditory speech

Figure 3B shows the relative distribution of activation for the two contrasts of interest ('visual speech – blank screen' in red and 'auditory speech – silence' in green). The overlap between red and green clusters is marked in yellow and this indicates where the common activation is located. However, a direct statistical analysis is required to make inferences about the probability of the common activation. Table 2 reports the results of a conjunction analysis. Visual speech and auditory speech predominantly involved common regions of the middle temporal gyrus bilaterally, likely to play a role in semantic identification (Friederici, 2002). It is interesting to note that both tasks reliably engaged semantic processing despite the fact that the difficulty of the number identification task and level of comprehension were not equally matched across visual and auditory speech conditions. Careful comparison of the conjoint activation with probabilistic maps of Heschl's gyrus (Penhune et al., 1997) and planum temporale (Westbury et al., 1999) revealed that common neural processing did not extend upwards from the middle temporal gyrus to include the primary auditory cortex nor any of the surrounding non-primary auditory fields. Nor did the activated region extend upwards and back enough to fall within the site for audio-visual integration identified in the superior temporal sulcus/posterior middle temporal gyrus (Beauchamp et al., 2004; Calvert et al., 2000). A second region of overlap occurred the medial superior frontal gyrus (BA 6) which could reasonably be involved in aspects of motoric planning, since both contrasts required a button press in the active task but none in the baseline task.

Neural bases for individual differences in speechreading ability

Correlating individual performance with cortical activation can help to clarify which brain regions gain access to language through speechreading. The results of the correlation analysis revealed brain regions where activation was a linear function of participants' ability to correctly identify the number in the visual speech sentences. Figure 4 illustrates the distribution of the speechreading activation for the two best-fitting voxels in the medial

superior frontal gyrus and the posterior cingulate cortex (negative and positive correlations respectively). Other brain regions to show similar patterns of correlation occurred in the frontal and occipital lobes and these are listed in Table 3. The distribution of the correlated activity is represented in Figure 3C where negative correlations are shown in cyan and positive correlations in pink. Identified regions were small in extent, but did reach the peak height probability threshold ($P < 0.001$). A negative correlation with speechreading has previously been shown in left inferior frontal gyrus corresponding to Broca's area (Paulesu et al., 1996), but the role of right inferior frontal gyrus (BA 47) is unclear. The peak in BA 6 was more anteriorly located than its counterpart in the subtraction analyses and could overlap with pre-SMA. Pre-SMA seems to be important in the procedural learning of new sequential movements (Hikosaka et al., 1996). Performance-related differences in motor activation might therefore reflect strategy differences in generating a button press, determined by either what has been speechread or at random. Positive correlations were also found in a number of brain regions. Bilateral clusters of activation occurred in the lingual gyrus of the visual cortex, and these were approximately 1 cm displaced from the midline. Their involvement suggests that low-level visual analysis is implicated in good speechreading. Additional regions implicated in good speechreading included the left fusiform gyrus, although this was slightly more medial to the basic visual speech network shown in Figure 3A, and the posterior cingulate cortex. Activation within the middle temporal gyrus, that was highly activated by visual speech and is associated with semantic comprehension, did not significantly increase with speechreading proficiency, but neither did it do so in other studies of speechreading ability (Paulesu et al., 2003). One speculative explanation for this null result is that perhaps the act of trying to understand speech generates a pattern of activation that is equivalent to successful comprehension.

The final analysis explored the extent to which speechreading elicited sound-based activation across individual subjects by using a mask to define the auditory cortex on the

superior temporal gyrus. The mask was determined using published anatomical criteria (details in the Methods section). A probability map of speechreading activation was computed by summing across individuals the occurrence of voxels within the mask that exceeded a threshold of $P<0.001$. Twenty eight of the 33 participants activated part of the superior temporal gyrus and, generally speaking, the activation was more extensive on the left than on the right (Figure 5). The maximum probability of activation at any specific voxel location reached 0.64, but this occurred at an inferior location within the mask where the probability of that voxel being within the superior temporal gyrus is below 25%. To determine whether sound-based activation varied with speechreading ability, the number of activated voxels within the mask was plotted as a function of each participant's speechreading score. There was a significant positive correlation in the left hemisphere (Pearson $r = 0.42$, $P < 0.02$), but not in the right (Pearson $r = -0.07$, ns). Hence, the data suggest that good speechreading engages the left superior temporal gyrus in a way that does not happen for poor speechreading.

Discussion

The many cognitive operations required to understand spoken sentences from a talker's face are mediated by an extensive network of brain activity. The different baseline conditions enabled us to decompose this network into components that could then be linked to different putative functional roles. The contrasts of the static face with the blank screen and of the gurning with the static face revealed that speechreading involves occipito-temporal regions that are associated more generally with the visual analysis of objects and scenes (Grill-Spector et al., 1998). The role of these visual regions in processing visual speech is uncertain because no significant visual activation was revealed here by the 'visual speech – gurning' contrast, nor in the 'lexical – nonlexical' contrast reported by Paulesu *et al.* (2003), yet we did observe increased activation in the left fusiform gyrus as a function of

speechreading proficiency. A number of visual speech studies have also reported visual activation over and above that evoked by non-linguistic facial movements in the vicinity of the inferoposterior temporal lobe (BA37/19) (Calvert et al., 1997; Calvert & Campbell 2003), but these two studies used a half-face presentation in which the mouth movements filled a large visual angle and were highly prominent. Factors such as the meaningfulness of the facial gestures for the task or their perceived visual salience might contribute to the individual differences in visual activation observed in the present study.

As expected, speechreading involved many brain areas, more extensively in the left than in the right hemisphere, that have also been implicated in language processing. Specifically these regions include the middle temporal gyrus, inferior and middle frontal gyri, and parietal lobe. There is evidence that the left inferior frontal and middle temporal gyri form part of a modality-independent brain system for language comprehension because they are equally activated by written and spoken words (Booth et al., 2002). Current models of speech perception and comprehension define both modality-independent and modality-specific components (e.g., Hickok & Poeppel, 2000; Bookheimer, 2002). Thus, there is value in looking towards these models to explore putative functional roles for our regions of speechreading activation. The involvement of the middle temporal gyrus, inferior and middle frontal gyri, and parietal lobe revealed in the present study is entirely consistent with the neural system proposed to underlie speech perception in the model of Hickok and Poeppel (2000). These authors propose two routes that both originate from the superior temporal gyrus (which encodes modality-specific, sound-based information from the auditory input) and then project to modality-independent modules. The two routes are differentially recruited according to the requirements of the speech task. Comprehension of speech primarily involves a region at the left temporo-parietal-occipital junction that provides an interface with distributed semantic representations. In contrast, tasks requiring access to sublexical segmental information involve a left fronto-parietal network. Hickok and Poeppel suggest that

the inferior frontal region codes speech in an articulatory context, including the sublexical segmental coding of speech and the subvocal rehearsal of phonological working memory, whereas the inferior parietal lobe interfaces between sound-based and articulatory representations of speech. Speechreading is likely to call upon both task modes because of the complex linguistic strategies required to meet the challenges of this task. Indeed many of our participants reported using a speechreading strategy that involved imagining the sound corresponding to the visual speech - a strategy that would strongly favour the fronto-parietal pathway. Rönnberg et al. (1998) propose that phonological processing is an essential prerequisite for speechreading because, operationally-speaking, an effective pattern recognition system cannot be based on linguistic facial movements alone, given the phonetic ambiguity of visual speech. Phonological processing is a necessary means of accessing lexical and semantic information during visual speech because it includes a recoding of the visual input that can be mapped onto a sound-based representation of the word and thereafter other components of the receptive language system. There is strong support for the notion that the left inferior frontal gyrus is not simply involved in the articulatory-based mechanisms of speech perception, but also has a key role in language *comprehension* (Bookheimer, 2002; Friederici, 2002). Evidence suggests that the left inferior frontal gyrus supports the formation of syntactic and semantic structure and syntactic working memory (Friederici, 2002), as well as the executive aspects of semantic processing (Bookheimer, 2002). Thus, the left inferior frontal gyrus might play multiple roles in comprehending sentences presented as visual speech.

Does the speechreading of syllable or word lists activate a similar neural network to that of sentences? Certainly, the extensive literature on auditory speech processing indicates that distributed regions in the left hemisphere contribute to semantic processing; with a left temporo-parietal region, that encompasses the superior and middle temporal gyri, being more highly activated by sentences than by word lists (see Narain et al., 2003). In the present study,

speechreading sentences activated this same region of the left hemisphere. A systematic comparison between the neural bases for speechreading word lists and sentences is complicated by the observation that the distribution of activation is also determined by other aspects of the speechreading task and by the type of baseline comparison. For example, the middle temporal gyrus seems more strongly recruited when the speech has a greater degree of linguistic complexity. Speechreading lists of digits between 1 and 10 did not engage the middle temporal gyrus (Calvert et al., 1997; Campbell et al., 2001; MacSweeney et al., 2000). Word lists engaged only the right middle temporal gyrus (Paulesu et al., 2003), while sentences engaged the same region on both sides (Ludman et al., 2000). The left inferior frontal gyrus is not strongly implicated in speechreading when the baseline condition controls for subvocal rehearsal such as counting or repeating words (Calvert et al., 1997; MacSweeney et al., 2000; 2002). Statistical contrasts that did not explicitly control for subvocal rehearsal did seem to engage the left inferior frontal gyrus, albeit more weakly than in the present study (syllables, Calvert & Campbell, 2003; words and pseudowords, Paulesu et al., 2003).

What process does auditory activation reflect?

Our stringent group analysis identified patterns of reliable visual-speech activation across the group of 33 participants, but this pattern did not include the primary auditory cortex on the medial two-thirds of Heschl's gyrus nor the non-primary auditory cortex on the surrounding superior temporal gyrus. In the group analysis, the only significant overlap between visual and auditory speech that occurred within the temporal lobe was in the middle temporal gyrus bilaterally. Nevertheless, individual analyses revealed that for some participants, the superior temporal gyrus *was* activated by visual speech. Hence, we provide the first evidence that the involvement of the auditory cortex in speechreading is dependent upon speechreading proficiency. Numerous neuroimaging studies have claimed that the visual perception of speech is supported by phonemic processing that occurs in the bilateral superior

temporal gyrus (Calvert et al., 1997; 2000; MacSweeney et al., 2000; 2002). The presence of direct associative links between the representations of visible speech-like articulations and their corresponding sounds certainly provides an attractive explanation of the data and is consistent with other demonstrations of the strong influence that one sensory system exerts on the other; both behaviourally (McGurk and MacDonald, 1976) and physiologically (Giraud et al., 2001; MacSweeney et al., 2002). Calvert et al. (1999) suggest that the auditory and visual systems could access one another via a connecting structure such as the claustrum. Rather than gaining direct, preferential access to the auditory system, an alternative interpretation is equally possible in our view; that visual speech might gain access to acoustic representations indirectly via the phonological recoding strategies that are supported by the inferior frontal gyrus. Neuroimaging studies have confirmed that inner speech critically depends on an interaction between the inferior frontal and superior temporal gyri (Shergill et al., 2002) and in the present study visual speech produced widespread activation in these regions of the left speech-dominant hemisphere. The available data cannot distinguish whether the superior temporal gyrus reflects direct activation by visual speech or indirect activation by inner speech. In future, measures of the timecourse of activation in the superior temporal gyrus could usefully contribute to this debate.

A widespread network of regions underlies individual differences in performance

The positive correlation between the extent of superior temporal gyrus activation and speechreading scores is certainly consistent with better speechreading being associated with a larger proportion of sentences that are correctly monitored. Experimental evidence from speechreading in hearing-impaired adults particularly identifies the importance of phonologically-based working memory processes for successful speechreading (Rönneberg, 1995; Rönneberg et al., 1998). A further neural correlate of the phonological processes involved in speechreading was described by Paulesu et al. (2003) where a high level of

activation in the left inferior frontal gyrus was associated with poor speechreading accuracy, suggested to reflect the greater effort that is required for phonological analysis when the task is difficult. Perhaps the reason for the lack of similar frontal changes in the present dataset is that the amount of cognitive effort required to speechread sentences, rather than word lists, is consistently high for all participants irrespective of their general competence. Although speechreading is generally challenging, the level of difficulty does vary from person to person. A region of the brain that has consistently been associated with task difficulty is the posterior cingulate cortex, with activity decreasing as the task places a greater load on stimulus discriminability and short-term memory (McKiernan et al., 2003). Taking individual error rates as an indicator of task difficulty, the current speechreading data are no exception to this rule. Hence, the changes in posterior cingulate might reflect a general reallocation of processing resources according to the difficulty of the speechreading task. Only one previous study has associated the posterior cingulate cortex with speechreading proficiency (MacSweeney et al., 2002). In this study deaf, but not hearing, participants showed posterior cingulate activity when speechreading was contrasted with gurning. Since deaf subjects perhaps find silent speechreading less difficult than normally-hearing subjects, a general processing capacity explanation would fit these results. However, MacSweeney et al. propose an alternative view that the posterior cingulate might play a task-specific role in speechreading in integrating contextual and stored information with on-line language processing (McGuire et al., 1999). Rönneberg's (1995) cognitive model of skilled speechreading also strongly implicates the ability to integrate prior context with incoming visual information as a prerequisite for skilled speechreading. Other involvement from visual processing areas in the posterior cortex is consistent with the prior claim from ERP data that perceptual and/or attentional mechanisms also underlie good speechreading (Shepherd et al., 1977; Shepherd, 1982). More recently, speechreading ability has been linked with the recruitment of early visual processing regions in cochlear implant users (Giraud et al., 2001).

Implantees engage visual association regions (BA18) when listening to meaningful sounds suggesting that sound input might raise an expectancy for concomitant complementary visual input. Certainly, when the auditory signal is acoustically degraded, speech comprehension would become more reliant on an enhanced coupling between speech sounds and their corresponding visible mouth movements. The precise detail of the audio-visual integration process still remains unclear, although the locations of activation in the present study would suggest that both low-level (retinotopic) and high-level (object-based) processes play a role.

In summary, the widespread pattern of speechreading activation for visually-presented sentences reflects the perceptual, linguistic, and motor response demands of the task. Fluent speechreading invokes multiple stages of the spoken language system, including inner speech. Skilled speechreading does not appear to be supported by a distinct biological system, but individual differences in performance are probably reflected in the efficiency of the links between sensory and higher-order systems; for example, in the integration of visual input with stored knowledge about articulatory gestures, their associated speech sounds and word meanings.

Methods

Subjects

Thirty three participants were selected from a pool of 56 screened volunteers so that their speechreading ability spanned a broad range. Speechreading ability was screened using 60 of the Bench-Kowel-Bamford (BKB) sentences (Bench and Bamford 1979) that had previously been validated by MacLeod and Summerfield (1987) to reliably assess the speechreading ability of normally-hearing adults. These short sentences spanned a range of speechreading difficulty, from very easy (e.g. “The small boy was asleep”) to very hard (e.g. “The grocer sells butter”).

The 33 participants had a mean age of 22 years and a range of 18 to 52 years and 22 were female. All participants were right-handed, were native speakers of British English, had normal hearing (pure tone thresholds <20dB HL between 250 and 4000 Hz inclusive), had either normal vision or vision corrected with contact lenses (none had 3 m Snellen chart ratings for visual acuity < 20/60) and had no history of neurological impairment. All participants gave informed written consent.

Stimuli

For the fMRI study, a set of 100 short sentences was taken from MacLeod and Summerfield (1990); the IHR Audio-visual Adaptive Sentence Lists. These sentences are sensitive to individual differences in speechreading skill, particularly among poorer speechreaders (MacLeod & Summerfield, 1990). All sentences contain three keywords and are neither very easy nor very hard to speechread, since when evaluated they were correctly speechread by fewer than half of Rosen and Corcoran's (1982) subjects. The sentences were modified so that on-line non-verbal measures of accuracy could be acquired during the fMRI study in order to avoid unnecessary head motion. A number, between one and ten, was incorporated into each sentence (e.g., "The *four* yellow leaves are falling") at varied positions. The 100 sentences are referred to as the IHR number sentences. For scoring, subjects were required to identify the number using a ten-button response box.

Both the screening sentences (Macleod & Summerfield, 1987) and the IHR number sentences were spoken by two talkers (the co-authors QS and DH). The head and shoulders of each talker were captured against a uniform green background (see Figure 1) and the talker maintained an expressionless, unblinking face. Spoken sentences were recorded onto mini digital video (DV) using a Canon MV1 digital camcorder and a Brüel and Kjær microphone (type 4165) and measuring amplifier (type 2636). Gurning and static faces were also captured on video. The gurning faces involved bilateral closed-mouth gestures, or twitches of the lower

face that were not clearly non-linguistic (Figure 1). Fifty different gurns were recorded for each talker. Static faces were freeze-frame clips of each talker's head and shoulders with an expressionless, closed mouth (Figure 1). Twenty of these clips were created for each talker, because the position of the talker's head varied slightly between sentences. All video clips were edited to *mpg* file format with a duration of 3.9 s in, and 0.25s onset and offset darkened ramps. For the auditory-speech condition, sound clips were the acoustical analogues of all the IHR number sentences, edited to *wav* file format.

Results from the screened volunteers confirmed that there was a strong positive correlation (Pearson $r = 0.86$, $P<0.01$) across individuals between their percentage of keywords correctly reported for the 60 BKB sentences and the percentage of numbers correctly identified for the 100 IHR number sentences. This result supports the idea that our fMRI test provided an appropriate surrogate on-line measure for the ability to extract meaningful linguistic information from the talking face.

fMRI task and protocol

Participants completed two fMRI tasks (the visual speech and auditory speech) in a single scanning session. To obtain a fair measure of speechreading ability, the auditory-speech task was always conducted after the visual-speech task because the same sentences were used for both conditions. In the visual-speech task, there were static face, gurning face, and visual speech conditions plus a blank-screen condition. All face conditions contained a sequence of ten 3.9 s video clips with a randomised, but equal, occurrence of either talker. Conditions were presented in a pseudo-randomised order. The auditory-speech task comprised only two experimental conditions; auditory speech and a silent baseline, presented in an alternating order. During the auditory task, a fixation cross was presented on the screen to constrain the focus of visual attention. Participants were provided with a two-handed, ten-button response box. In the visual-speech and auditory-speech conditions, participants pressed

the button that corresponded to the specific number that occurred in each IHR number sentence. For the static- and gurning-face conditions, they were instructed to respond to each clip by pressing the buttons in a numerical sequence (one to ten). No response was required in the blank-screen or silent baseline conditions. Button presses were logged for off-line analysis. The total duration for all epochs was 39 s and there were 10 epochs for all experimental conditions. Thus, the visual-speech task lasted 26 minutes, while the auditory-speech task lasted 13 minutes.

Participants lay supine in the scanner and wore a set of prismatic goggles to enable them to see a projector screen that was 3 m away at the end of the scanner bed. The projection of the talker's face gave a visual angle of 18.4˚. Participants also wore a custom-built MR-compatible electrostatic headset (http://www.ihr.mrc.ac.uk/research/technical/index.php) for the presentation of the auditory speech and for 30 dB attenuation of the background acoustic gradient noise.

Scanning was performed on a 3 T MR scanner using a custom-built head volume coil (Bowtell and Peters, 1999). Hardware limitations imposed the use of coronal slices which were acquired in a sequential order from the front to the back of the head. Sets of T2*-weighted EPI data (TR = 9750 ms, TE = 36 ms) consisted of 38 contiguous coronal slices with a voxel size of 4 mm^3 and a matrix size of 128 x 64 elements. The sparse imaging protocol (Hall et al., 1999) was employed to reduce the interference of the intense background noise on the pattern of speech-sound activation. Each set of images was acquired in 2660 ms, leaving 7090 ms of background quiet in between each volume acquisition. Since each stimulus condition was 39 s in duration, 4 sets of images were acquired in each stimulus condition. We acquired 160 scans during the visual speech task and 80 scans during the auditory speech task. Often the 38 coronal slices excluded the cortical frontal pole from the field of view and so an additional whole-brain set of 64 coronal EPI slices was acquired to facilitate post-processing.

Image analysis

Images were analysed using SPM99 (http://www.fil.ion.ucl.ac.uk/spm). A realignment correction was applied to remove head movements in both translation and rotation dimensions. To closely align the two EPI datasets for each individual, the time series for the visual-speech task was realigned to the last scan in that series and the time series for the auditory-speech task was realigned to the first scan in that series because these two scans occurred most closely together in time. The 64-slice EPI was co-registered to the mean of the realigned images using an algorithm based on mutual information (Maes et al., 1997). The following step involved computing the linear and non-linear transformations needed to map the EPI data into standard brain space for averaging the different brains. To do this, we first computed the spatial normalisation parameters for the 64-slice EPI by matching it to the SPM99 EPI template. All EPI data showed a gradient of signal decrease in the cerebellum because of regional loss of sensitivity in the volume coil and so, to improve the match between our data and the EPI template we applied a similar gradient to the template. The computed normalisation parameters were then applied to both EPI time series and normalised images were written, maintaining the original voxel size of 4 mm^3. This procedure generally optimized the match between our data and standard brain space. However in the cerebellum, we took care to localize activation with respect to our normalised group-averaged 64-slice EPI because this region remained slightly stretched downwards. The final post-processing step was to spatially smooth the data using a gaussian kernel of 8 mm full width at half maximum.

For each participant, stimulus effects for the visual-speech and auditory-speech tasks were modelled separately at the first-level using a fixed-effects analysis that computed the within-subject error variance (residual scan-to-scan variability) as the only variance component. General linear models comprised one term for each condition that was a simple vector of 1 and 0s to define the stimulus on/off periods. Thus, there were four variables in the visual-speech model and two variables in the auditory speech model. We made the

assumption, common to the modelling of most blocked designs, that the haemodynamic response reached a plateau approximately 6 s after onset and remained constant for the duration of the stimulus. Low-frequency physiological noise and scanner drift was modelled by high-pass filtering the visual-speech time series at 0.1 cycles/minute and the auditory-speech time series at 0.4 cycles/minute.

To decompose the speechreading network into different functional components, key pairwise statistical contrasts between variables were specified; i) static face versus blank screen, ii) gurning face versus static face, iii) visual speech versus gurning face, and iv) auditory speech versus silence. The outputs of these individual analyses were then entered into four second level, random-effects (RFX) analyses which tested the significance of the pairwise contrast across the group by assessing the between-subject variance component. To statistically evaluate the common pattern of activity shared by visual speech and auditory speech tasks, a further RFX analysis directly compared these two contrasts by testing for significant effects in *both* speech contrasts. A valid conjunction inference about the prevalence of the common effect used the conjunction null hypothesis, implemented in SPM2 (Brett et al., 2004). All results were thresholded at $P < 0.05$, with a correction for multiple comparisons across the whole brain. This type of analysis provides a conservative estimate about the typical patterns of activation (Holmes & Friston, 1998).

Two different statistical approaches were employed to investigate the nature of any brain activation that was related to speechreading ability. Both analyses were based on the pairwise contrast between the visual speech and the blank screen conditions to include the entire network of areas involved at all stages of the speechreading process. The first approach used a second-level regression analysis of the contrast images to determine whether the magnitude of the activation signal between subjects varied as a linear function of their speechreading scores on the IHR number task. Results were thresholded using a voxel height threshold of $P < 0.001$. The second approach used the individual thresholded activation maps

to investigate whether the extent of significant activation in the auditory cortex was dependent upon speechreading ability. The chosen height threshold was again $P < 0.001$, but also applying a correction for the volume of the superior temporal gyrus. This volume was obtained by the summation of maps for the outermost anatomical boundaries of Heschl's gyrus (Penhune et al., 1997) and planum temporale (Westbury et al., 1999). The volume was then used as a mask to extract the number of supra-threshold voxels within the region of interest for each participant.

The location of activation is reported using anatomical and Brodmann's labels. These were obtained by transforming the voxel coordinates into the Talairach coordinate space (Brett et al., 2001) and referencing to the atlas of Talairach and Tournoux (1988).

Acknowledgements This research was supported by the MRC, including an MRC research studentship G78/6969 awarded to CF. The authors thank Mr John Foster and Miss Kay Head for providing scanning support.

References

Bamford, J., & Wilson, I. (1979). Methodological and practical aspects of the BKB Sentence lists. In Bench, J. and Bamford, J. (Eds). Speech-hearing tests and the spoken language of hearing-impaired children. Academic Press, London.

Beauchamp, M.S., Lee, K.E., Argall, B.D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron, 41, 809-823.

Bernstein, L.E., Auer, E.T., Moore, J.K., Ponton, C.W., Don, M. & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. Cognitive Neuroscience, 13, 311-315.

Binnie, C.A. (1977). Attitude changes following speechreading training. Scandanavian Audiology, 6, 13-19.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. Annual Review of Neuroscience, 25, 151–188.

Booth, JR., Burman, D.B., Meyer, J.R., Gitelman, D.R., Parrish, T.B., & Mesulam, M.M. (2002). Modality independence of word comprehension. Human Brain Mapping, 16, 251–261.

Bowtell, R., Mansfield, P., Coxon, R.J., Harvey, P.R. & Glover, P.M. (1994). High-resolution echo-planar imaging at 3.0 T. Magnetic Resonance in Material Physics, Medicine and Biology, 2, 1-5.

Brett, M., Nichols, T., Andersson, J., Wager, T. & Poline, J-B. (2004). When is a conjunction not a conjunction? 10th Annual Conference of the Organisation for Human Brain Mapping, July, Budapest.

Brett, M., Christoff, K., Cusack, R. & Lancaster, J.L. (2001). Using the Talairach atlas with the MNI template. NeuroImage, 13 (6), S85.

Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iverson, S.D., & David, A.S. (1999). Response amplification in sensory-specific cortices during cross modal binding. NeuroReport, 10, 2619-2623.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P.K., Woodruff, P.W.R., Iversen, S.D. & David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science, 276, 593-596.

Calvert, G.A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. Journal of Cognitive Neuroscience, 15, 57-70.

Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology, 10, 649-657.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M. & David, A.S. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Cognitive Brain Research, 12, 233-243.

Decety, J. & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. Trends in Cognitive Sciences, 3, 172-178.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences, 6, 78-84.

Friederici, A., Meyer, M. & Von Cramon, D.Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain and Language, 74, 289-300.

Giraud, A-L., Price, C.J., Graham, J.M., Truy, E. & Frackowiak, R.S.J. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation, Neuron, 30, 657-664.

Grill-Spector, K., Knouf, N. & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category perception. Nature Neuroscience, 7, 555-562.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach R. (1998).Cue invariant activation in object-related areas of the human occipital lobe. Neuron, 21, 191–202.

Hall, D,A,, Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R. Gurney, E. & Bowtell, R.W. (1999). Sparse temporal sampling in auditory fMRI. Human Brain Mapping, 7, 213-223.

Hasnain, M.K., Fox, P.T. & Woldorff, M.G. (1998). Intersubject variability of functional areas in the human visual cortex. Human Brain Mapping, 6, 301-315.

Heider, F & Heider, G. (1940). An experimental investigation of lipreading. Psychological Monographs, 52, 124-153.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences, 4, 131-138.

Hikosaka, O., Sakai, K., Miyauchi, S., Takino, R., Sasaki, Y., & Pütz, B. (1996). Activation of human presupplementary motor area in learning of sequential procedures: a functional MRI study. Journal of Neurophysiology, 76, 617-621.

Holmes, A.P., & Friston, K.J. (1998). Generalisability, random-effects and population inference. NeuroImage 7, S754.

Jeffers, J., & Barley, M. (1971). Speechreading (lipreading). Springfield, IL: Thomas.

Ludman, C.N., Summerfield, A.Q., Hall, D., Elliott, M., Foster, J., Hykin, J.L., Bowtell, R. & Morris, P.G. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. British Journal of Audiology, 34, 225-230.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution to speech perception in noise. British Journal of Audiology, 21, 131-141.

Macleod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. British Journal of Audiology, 24, 29-43.

MacSweeney, M., Amaro, E., Calvert, G.A., Campbell, R., David, A.S., McGuire, P., Williams, S.C.R., Woll, B. & Brammer, M.J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. NeuroReport, 11, 1729-1733.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. & Suetens, P. (1997). Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 16, 187-198.

Maguire, E.A., Frith, C.D. & Morris, R.G.M. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. Brain, 122, 1839–1850.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

McKiernan, K.A., Kaufman, J.N., Kucera-Thompson, J & Binder, J.R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. Journal of Cognitive Neuroscience, 15, 394-408.

Narain, C., Scott, S.K., Wise, R.J.S., Rosen, S., Leff, A., Iversen, S.D., & Matthews, P.M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. Cerebral Cortex, 13, 1362-1368.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N.A., De Giovanni, V., Sensolo, S. & Fazio, F. (2003). A functional-anatomical model for lipreading. Journal of Neurophysiology, 90, 2005-2013.

Penhune, V.B., R.J. Zatorre, R.J., Macdonald, J.D. & Evans, A.C. (1996). Interhemispheric Anatomical Differences in Human Primary Auditory Cortex: Probabilistic Mapping and Volume Measurement from Magnetic Resonance Scans. *Cerebral Cortex* 6: 661-672.

Rönneburg, J., Arlington, S., Lyxell, B. & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. Journal of Speech and Hearing Research, 32, 725-735.

Rönneberg, J. (1995). What makes a skilled speechreader? In G. Plant, & K-E Spens (Eds.), Profound Deafness and Speech Communication (pp. 393-416). San Diego, CA: Singular Publishing Group Inc.

Rönnberg, J., Samuelsson, S. & Lyxell, B. (1998). Conceptual constraints in sentence-based lipreading in the hearing impaired. In R. Campbell, B. Dodd & D. Burnham (Eds.), Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech (pp. 143-153). Hove, UK: Psychology Press.

Rosen, S.M., & Corcoran, T. (1982). A video-recorded test of lip-reading for British English. British Journal of Audiology, 16, 245-254.

Samar, V.J., & Sims, D.G. (1983). Visual evoked-response correlates of speechreading performance in normal-hearing adults: A replication and factor analytic extension. Journal of Speech and Hearing, 26, 2-9.

Samar, V.J., & Sims, D.G. (1984). Visual evoked-response components related to speechreading and spatial skills in hearing and hearing-impaired adults. Journal of Speech and Hearing Research, 27, 162-172.

Shepherd, D.C., DeLavergne, R.W., Frueh, F.X., & Clobridge, C. (1977). Visual-neural correlate of speechreading ability in normal-hearing adults. Journal of Speech and Hearing, 20, 752-765.

Shepherd, D.C. (1982). Visual-neural correlate of speechreading ability in normal-hearing adults: Reliability. Journal of Speech and Hearing Research, 25, 521-527.

Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M. & McGuire, P.K. (2002). Modulation of activity in temporal cortex during generation of inner speech. Human Brain Mapping, 16, 219-227.

Sumby, W.H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. Journal of the Acoustical Society of America, 26, 212-215.

Summerfield, A.Q. (1991). Visual perception of phonetic gestures. In, I.G. Mattingly and M. Studdert-Kennedy (Eds.), Modularity and the Motor Theory of Speech Perception. Lawrence Erbaum Associates: New Jersey. pp. 117-137.

Summerfield, A.Q. (1987). Some preliminaries to a comprehensive account of audiovisual speech perception. In B. Dodd & R. Campbell (Eds.), Hearing by eye: The psychology of lipreading (pp. 3-51). Hillsdale, NJ: Erlbaum.

Talairach, J. & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. Thieme: Stuttgart.

Tanji, J., Shima, K. & Mushiake, H. (1996). Multiple cortical motor areas and temporal sequencing of movements Cognitive Brain Research, 5, 117-122.

Watkins, K.E., Strafella, A.P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involed in speech production. Neuropsychologia, 41, 989-994.

Westbury, C.F., Zatorre, R.J. & Evans, A.C. (1999). Quantifying variability in the planum temporale: A probability map. Cerebral Cortex, 9, 392-405.

Figure legends

Figure 1. Freeze-frame clips of one of the talkers (QS) used in the present study. Both the gurning face and visual speech clips were extracted from a 3.9 s recording of dynamic facial articulations.

Figure 2. Histogram showing the distribution of speechreading scores across the group of 33 participants. Scores reflect the proportion of numbers that were correctly identified in the IHR number sentences during the scanning experiment.

Figure 3. General summary views of the distributed brain activity in the different task contrasts overlaid onto lateral, superior and posterior surface views of the mean normalised group brain. Both lateral and medial activations are projected onto the cortical surface. Row (A) illustrates the hierarchy of the speechreading network that was determined by the different cognitive contrasts. Green denotes those regions engaged by processing the static face, blue denotes those regions involved in the analysis of non-linguistic facial movements and red denotes regions activated by visual speech. The regions in yellow in Row (B) represent the common processing circuits for visual (red) and auditory (green) speech. Activation that was dependent upon speechreading ability is shown in row (C). Regions in pink have a greater level of activation when speechreading is more skilled, regions in cyan have a lower level.

Figure 4. The speechreading skill of individual participants exerted both positive and negative effects on the level of brain activation in different brain regions. In the medial superior frontal gyrus activity decreases with better speechreading skill, whereas in the posterior cingulate cortex activity increases with better speechreading skill. Closed circles represent the adjusted activation signal for each subject and the line represents the best-fitting linear regression.

Figure 5. Although group-averaged activity within the superior temporal gyrus did not reach significance, there was some individual activation in this region at $P < 0.001$. The coloured region is a functional probability map showing where and in what percentage of participants significant activation occurred.

Figure 6. A graph of the data presented in Figure 5 showing the extent of superior temporal gyrus activation as a function of speechreading ability. Triangles show data for the left hemisphere and circles show data for the right hemisphere. Only the left superior temporal gyrus shows a significant linear relationship.

Table 1. Areas showing the systematic recruitment of additional brain regions areas for processing static, dynamic and linguistically-meaningful facial movements and their commonality with processing auditory speech. Co-ordinates are reported in the standard MNI brain space and refer to voxels that are 4 mm^3 in size. Z values are reported for the peak voxel within each cluster where the height threshold exceeds a level of $P < 0.05$, corrected for multiple comparisons across the whole brain. BA labels approximate to the corresponding Brodmann anatomical reference scheme.

Table 2. Brain activation that is commonly activated both by visual and auditory speech. Brain regions survive a height threshold of $P < 0.05$, corrected for multiple comparisons across the whole brain.

Table 3. Pattern of activation that is a function of speechreading ability. Brain regions survive an uncorrected height threshold of $P < 0.001$.

Peak MNI coordinate (mm)			Cluster size in voxels	Max Z value	Side	Anatomical region	BA
x	y	z					
Static face – blank screen							
0	-4	52	68	6.23	Mid	Medial superior frontal gyrus	6
-20	-92	-36	43	5.86	L	Fusiform gyrus / Cerebellum	18,19 / -
32	-80	-40	34	5.17	R	Fusiform gyrus / Cerebellum	18,19 / -
-40	-20	60	146	5.78	L	Precentral gyrus	4
Gurning - static faces							
-48	80	-20	54	5.67	L	Middle occipital gyrus	19
48	-68	-16	16	5.25	R	Middle occipital gyrus	19
Visual speech– gurning							
0	8	60	292	7.52	Mid	Medial superior frontal gyrus	6
-48	20	20	523	7.24	L	Inferior frontal gyrus / Middle frontal gyrus / Precentral gyrus	44,45 / 9 / 6
40	24	-8	76	6.46	R	Inferior frontal gyrus	44,47
48	12	36	81	6.16	R	Middle frontal gyrus	9
-60	-44	-4	178	6.45	L	Middle temporal gyrus	21
64	-28	-12	89	6.80	R	Middle temporal gyrus	21
-40	-60	40	96	6.00	L	Inferior parietal lobe	7
-8	-24	4	29	5.63	L	Thalamus	-
Auditory speech – silence							
0	4	52	39	6.09	Mid	Medial superior frontal gyrus	6
-48	-4	48	37	5.76	L	Inferior frontal gyrus / Precentral gyrus	9 / 6
-32	-8	64	10	5.27	L	Middle frontal gyrus	6
-44	-32	4	572	7.80	L	Superior temporal gyrus / Middle temporal gyrus	41, 42, 22 / 21
64	-28	-8	437	Inf	R	Superior temporal gyrus / Middle temporal gyrus	41, 42, 22 / 21
-52	-40	48	65	5.73	L	Inferior parietal lobe	40

Table 1. Areas showing the systematic recruitment of additional brain regions areas for processing static, dynamic and linguistically-meaningful facial movements and their commonality with processing heard speech. Co-ordinates are reported in the standard MNI brain space. T values are reported for the peak voxel within each cluster where the height threshold exceeds a level of $P < 0.05$, corrected for multiple comparisons across the whole brain. BA labels approximate to the corresponding Brodmann anatomical reference scheme.

Peak MNI coordinate (mm)			Cluster size in voxels	Max Z value	Side	Anatomical region	BA
x	y	z					
Visual speech – blank screen AND Auditory speech – silence							
0	4	52	41	7.04	Mid	Medial superior frontal gyrus	6
-36	-8	60	13	5.64	L	Middle frontal gyrus	6
60	-40	-16	70	5.34	R	Middle temporal gyrus	21
-60	-48	0	142	7.32	L	Middle temporal gyrus	21
-52	-40	48	68	6.17	L	Inferior parietal lobe	40

Table 2. Brain activation that is commonly activated both by visual and auditory speech. Brain regions survive a height threshold of $P < 0.05$, corrected for multiple comparisons across the whole brain.

Peak MNI coordinate (mm)			Cluster size in voxels	Max Z value	Side	Anatomical region	BA
x	y	z					
Negative linear correlation with speechreading ability							
4	28	60	4	3.68	R	Medial superior frontal gyrus	6
20	24	60	6	3.55	R	Medial superior frontal gyrus	6
36	12	-16	4	3.49	R	Inferior frontal gyrus	47
Positive linear correlation with speechreading ability							
-16	-88	-4	4	3.39	L	Lingual gyrus	17, 18
12	-80	0	9	3.36	R	Lingual gyrus	17, 18
-20	-52	-12	11	3.67	L	Fusiform gyrus	37
4	-28	48	5	3.76	R	Posterior cingulate cortex	31

Table 3. Pattern of activation that is a function of speechreading ability. Brain regions survive an uncorrected height threshold of $P < 0.001$.

Peak MNI coordinate (mm)			Cluster size in voxels	Max T value	Side	Anatomical region	BA
x	y	z					
Static face – blank screen							
0	-4	52	68	8.81	Mid	Medial superior frontal gyrus	6
-20	-92	-36	43	7.94	L	Fusiform gyrus / Cerebellum	18,19 / -
32	-80	-40	34	6.53	R	Fusiform gyrus / Cerebellum	18,19 / -
-40	-20	60	146	7.76	L	Precentral gyrus	4
Gurning - static faces							
-48	80	-20	54	7.52	L	Middle occipital gyrus	19
48	-68	-16	16	6.68	R	Middle occipital gyrus	19
Visual speech– gurning							
0	8	60	292	12.66	Mid	Medial superior frontal gyrus	6
-48	20	20	523	11.67	L	Inferior frontal gyrus / Middle frontal gyrus / Precentral gyrus	44,45 / 9 / 6
40	24	-8	76	9.39	R	Inferior frontal gyrus	44,47
48	12	36	81	8.64	R	Middle frontal gyrus	9
-60	-44	-4	178	9.35	L	Middle temporal gyrus	21
64	-28	-12	89	10.33	R	Middle temporal gyrus / STS	21 / -
-40	-60	40	96	8.26	L	Inferior parietal lobe	7
-8	-24	4	29	7.45	L	Thalamus	-
Auditory speech – silence							
0	4	52	39	8.46	Mid	Medial superior frontal gyrus	6
-48	-4	48	37	7.72	L	Inferior frontal gyrus / Precentral gyrus	9 / 6
-32	-8	64	10	6.72	L	Middle frontal gyrus	6
-44	-32	4	572	13.74	L	Superior temporal gyrus / Middle temporal gyrus	41, 42, 22 / 21
64	-28	-8	437	15.90	R	Superior temporal gyrus / Middle temporal gyrus	41, 42, 22 / 21
-52	-40	48	65	7.65	L	Inferior parietal lobe	40

Table 1. Areas showing the systematic recruitment of additional brain regions areas for processing static, dynamic and linguistically-meaningful facial movements and their commonality with processing heard speech. Co-ordinates are reported in the standard MNI brain space. T values are reported for the peak voxel within each cluster where the height threshold exceeds a level of $P < 0.05$, corrected for multiple comparisons across the whole brain. BA labels approximate to the corresponding Brodmann anatomical reference scheme.

Peak MNI coordinate (mm)			Cluster size in voxels	Max T value	Side	Anatomical region	BA
x	y	z					
Visual speech – blank screen AND Auditory speech – silence							
0	4	52	41	8.70	Mid	Medial superior frontal gyrus	6
36	-8	60	13	6.45	R	Middle frontal gyrus	6
-60	-40	-16	70	6.02	L	Middle temporal gyrus	21
60	-48	0	142	9.20	R	Middle temporal gyrus	21
52	-40	48	68	8.70	R	Inferior parietal lobe	40

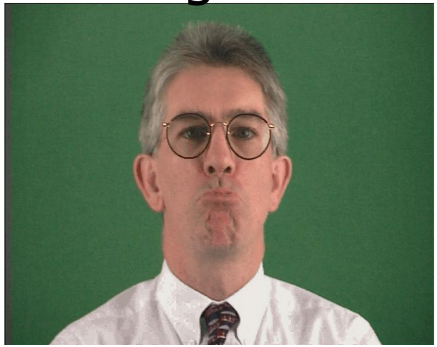
Table 2. Brain activation that is commonly activated both by visual and auditory speech. Brain regions survive a height threshold of $P < 0.05$, corrected for multiple comparisons across the whole brain.

Peak MNI coordinate (mm)			Cluster size in voxels	Max T value	Side	Anatomical region	BA
x	y	z					
Negative linear correlation with speechreading ability							
4	2	60	4	4.15	R	Medial superior frontal gyrus	6
20	24	60	6	3.98	R	Medial superior frontal gyrus	6
36	12	-16	4	3.90	R	Inferior frontal gyrus	47
Positive linear correlation with speechreading ability							
-16	-88	-4	4	3.76	L	Lingual gyrus	17, 18
12	-80	0	9	3.73	R	Lingual gyrus	17, 18
-20	-52	-12	11	4.15	L	Fusiform gyrus	37
4	-28	48	5	4.27	R	Posterior cingulate cortex	31

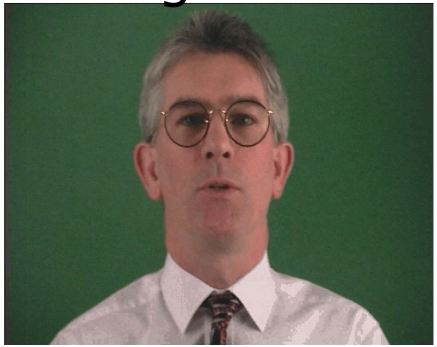
Table 3. Pattern of activation that is a function of speechreading ability. Brain regions survive an uncorrected height threshold of $P < 0.001$.

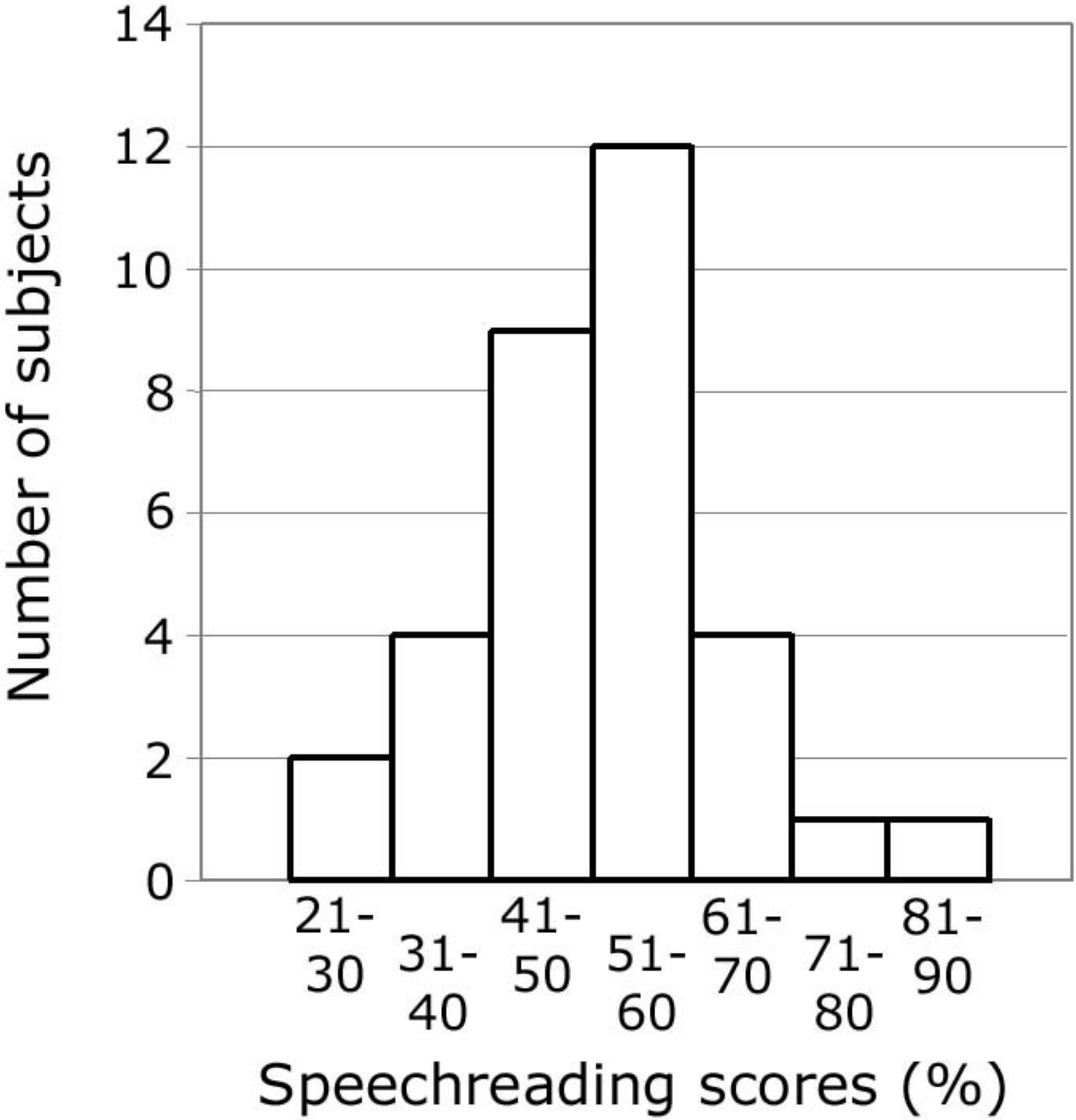
Static face

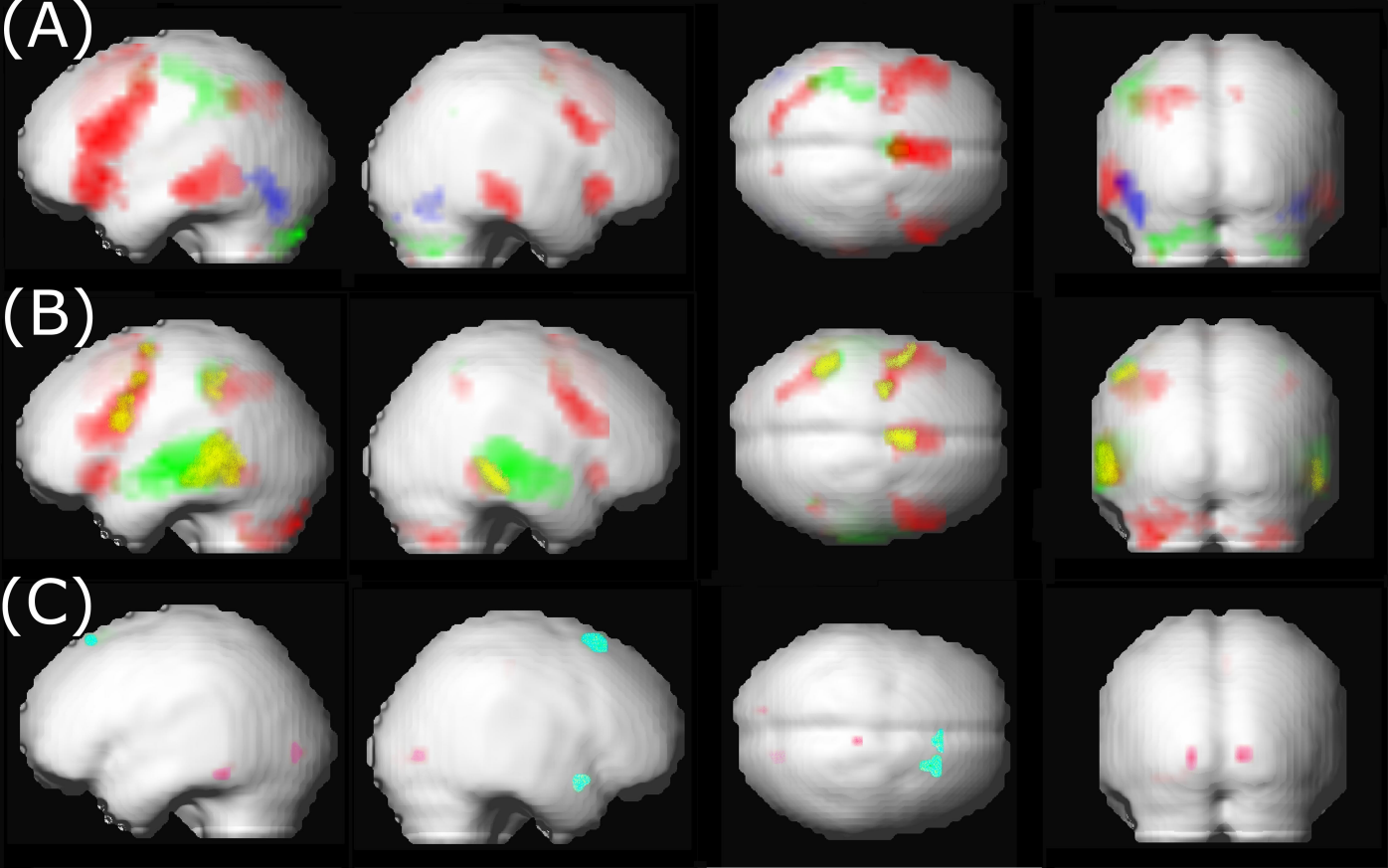
Gurning face

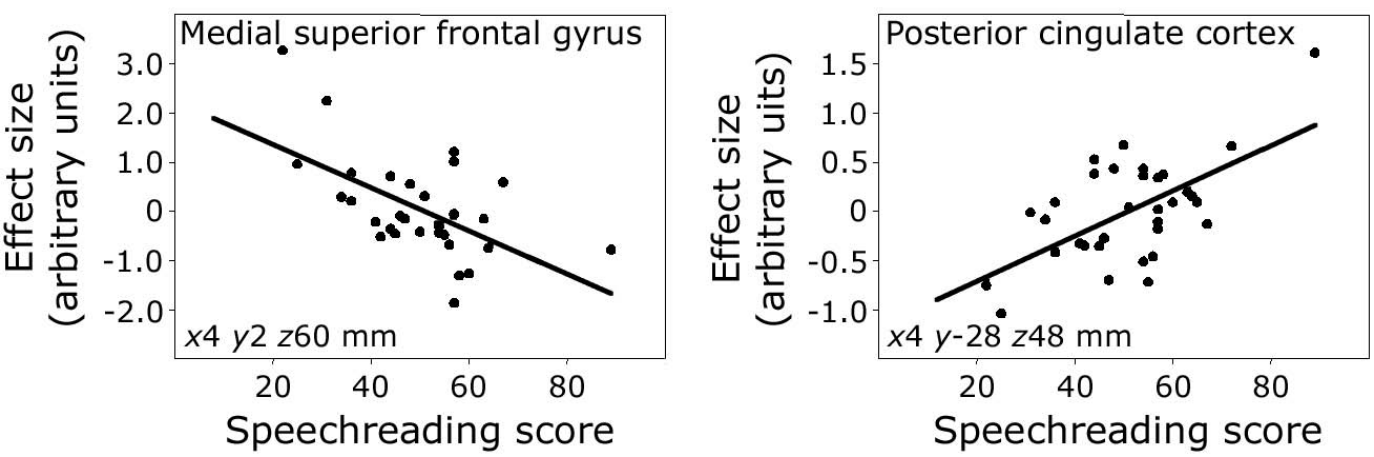


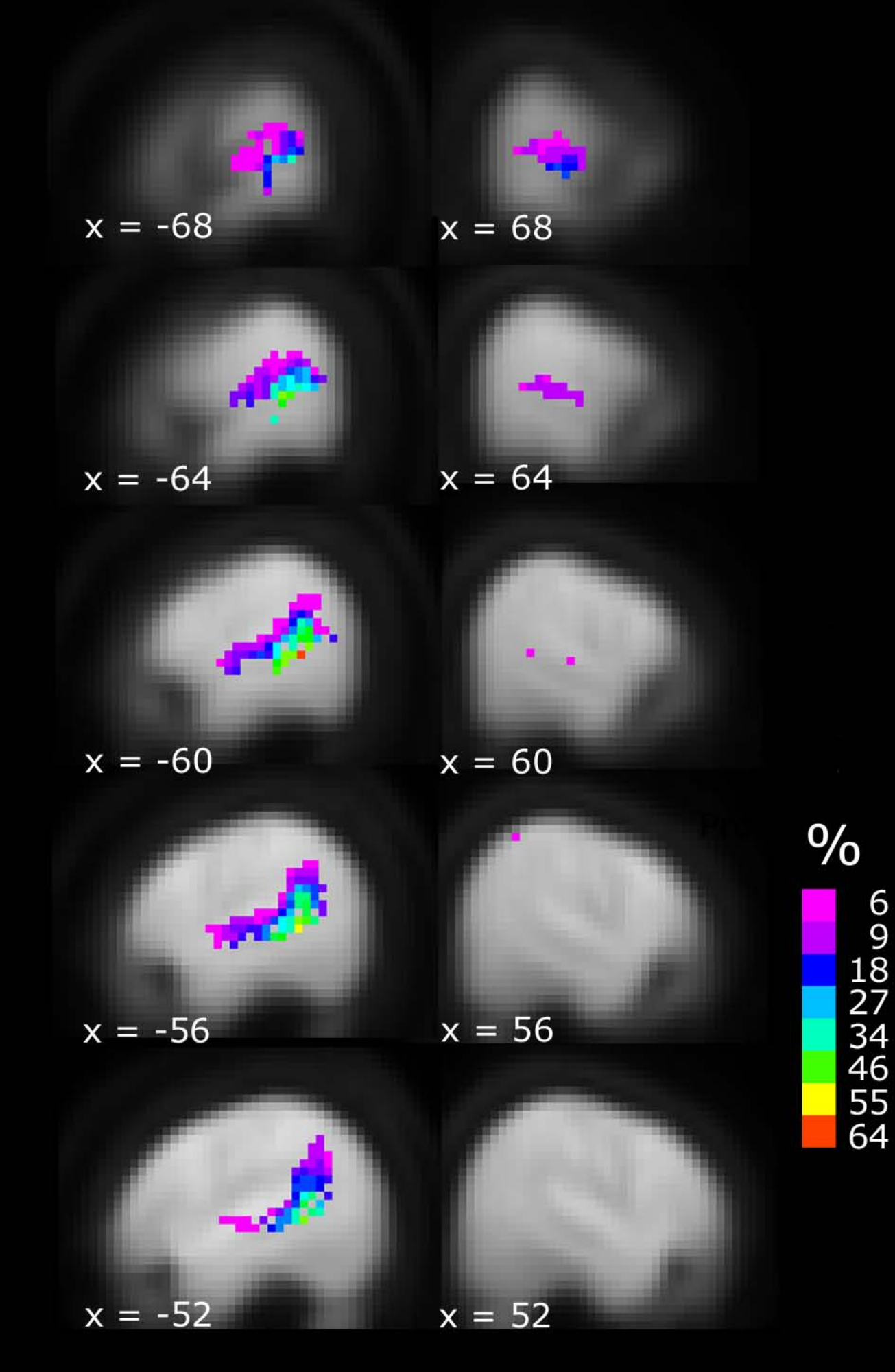
Talking face

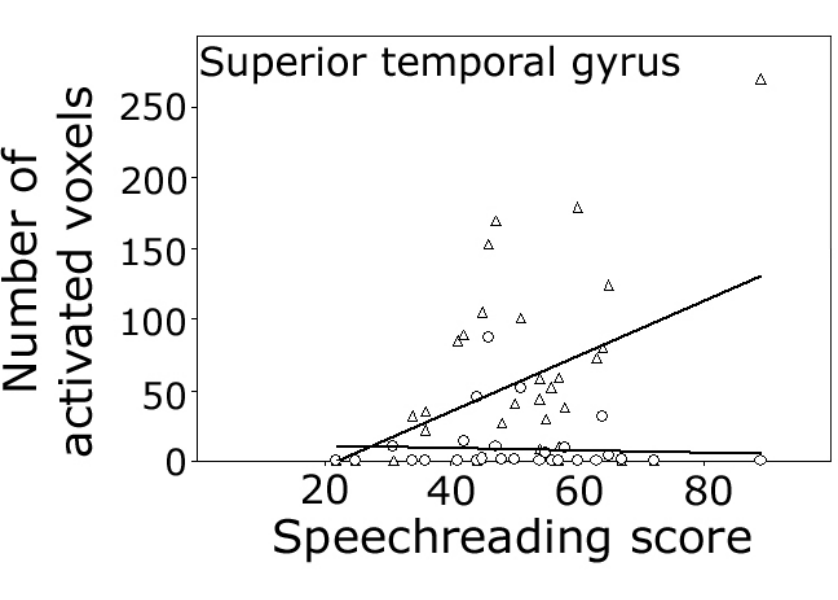



14
12
10
8
6
4
2
0
Number of subjects
21-30
31-40
41-50
51-60
61-70
71-80
81-90
Speechreading scores (%)


(A)
(B)
(C)


Medial superior frontal gyrus
Effect size (arbitrary units)
3.0
2.0
1.0
0
-1.0
-2.0
x4 y2 z60 mm
20
40
60
80
Speechreading score
Posterior cingulate cortex
Effect size (arbitrary uits)
1.5
1.0
0.5
0
-0.5
-1.0
x4 y-28 z48 mm
20
40
60
80
Speechreading score


x = -68
x = 68
x = -64
x = 64
x = -60
x = 60
x = -56
x = 56
x = -52
x = 52
%
6
9
18
27
34
46
55
64


Superior temporal gyrus
Number of activated voxels
250
200
150
100
50
0
20
40
60
80
Speechreading score