

Highlights

- Talker gaze is an important social cue during speech comprehension
- Neural responses to gaze were measured during perception of degraded sentences
- Gaze direction modulated activation in left-lateralized superior temporal cortex
- Left lateralization became stronger when speech was less intelligible
- Results suggest task-dependent flexibility in cortical responses to gaze

RUNNING HEAD: Neural responses to talker gaze

***You talkin' to me?* Communicative talker gaze activates left-lateralized superior temporal cortex during perception of degraded speech.**

**Carolyn McGettigan^{1,2}, Kyle Jasmin², Frank Eisner^{2,3}, Zarinah K Agnew^{2,4},
Oliver J Josephs^{2,5}, Andrew J Calder⁶, Rosemary Jessop², Rebecca P Lawson^{2,5},
Mona Spielmann², Sophie K Scott²**

1 Department of Psychology, Royal Holloway University of London, Egham Hill,
Egham TW20 0EX, UK

2 Institute of Cognitive Neuroscience, University College London, 17 Queen Square,
London WC1N 3AR, UK

3 Donders Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen,
Netherlands

4 Department of Otolaryngology, University of California, San Francisco, 513
Parnassus Avenue, San Francisco, CA.

5 Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University
College London, 12 Queen Square, London WC1N 3BG, UK.

6 MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF,
UK

Corresponding author:

Carolyn McGettigan, Department of Psychology, Royal Holloway, University of
London, Egham Hill, Egham TW20 0EX, UK

Email: Carolyn.McGettigan@rhul.ac.uk

Abstract

1
2 Neuroimaging studies of speech perception have consistently indicated a left-
3
4 hemisphere dominance in the temporal lobes' responses to intelligible auditory speech
5
6 signals (McGettigan & Scott, 2012). However, there are important communicative
7
8 cues that cannot be extracted from auditory signals alone, including the direction of
9
10 the talker's gaze. Previous work has implicated the superior temporal cortices in
11
12 processing gaze direction, with evidence for predominantly right-lateralized responses
13
14 (Carlin & Calder, 2013). The aim of the current study was to investigate whether the
15
16 lateralization of responses to talker gaze differs in an auditory communicative
17
18 context. Participants in a functional MRI experiment watched and listened to videos
19
20 of spoken sentences in which the auditory intelligibility and talker gaze direction were
21
22 manipulated factorially. We observed a left-dominant temporal lobe sensitivity to the
23
24 talker's gaze direction, in which the left anterior superior temporal sulcus/gyrus and
25
26 temporal pole showed an enhanced response to direct gaze – further investigation
27
28 revealed that this pattern of lateralization was modulated by auditory intelligibility.
29
30 Our results suggest flexibility in the distribution of neural responses to social cues in
31
32 the face within the context of a challenging speech perception task.
33
34
35
36
37
38
39
40
41
42
43

Keywords: Gaze, fMRI, speech comprehension, laterality

Introduction

1
2 Spoken communication can only be described in part by reference to the exchange of
3 linguistic messages. Natural conversation often occurs face-to-face, where
4 interlocutors display facial expressions, gestures and non-verbal vocalizations (such
5 as laughter) in order to enhance understanding, and to signal social cues such as
6 mood, affiliation and intent. However, to date, relatively little is known about how the
7 brain processes social and linguistic cues within the same communicative context.
8
9

10
11
12
13
14
15
16
17
18
19 Neuroimaging studies of auditory speech intelligibility in the healthy adult brain have
20 attempted to isolate the neural responses to intelligible (or partially intelligible)
21 speech signals by comparison with acoustically complex, unintelligible control
22 conditions (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Evans et al., 2014;
23 Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000) and by using parametric
24 modulations of speech intelligibility, for example by varying the number of channels
25 in noise-vocoded speech (Davis & Johnsrude, 2003; McGettigan et al., 2012; Obleser,
26 Wise, Alex Dresner, & Scott, 2007; Scott, Rosen, Lang, & Wise, 2006). This work
27 has identified that the process of extracting an intelligible message from an auditory
28 speech signal engages an anterior-going pathway in the superior temporal lobes
29 (Evans et al., 2014; Scott et al., 2000) as well as responses in the inferior frontal gyrus
30 (IFG), anterior insula and premotor cortex (Adank, 2012a; Davis & Johnsrude, 2003;
31 Eisner et al., 2010; Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012;
32 McGettigan et al., 2012). Speech comprehension can also be manipulated
33 experimentally through alternate methods, such as the comparison of words with
34 pseudowords, and the use of semantic and syntactic violations, revealing similar loci
35 (see Adank, 2012b). Although some authors argue that the perceptual processing of
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 speech is bilateral in the temporal lobes (Hickok & Poeppel, 2007; Okada et al., 2010)
2 our work has strongly suggested a left hemisphere dominance for intelligible speech
3 perception (including perception of sentences, words, syllables, phonemes, syntactic
4 and semantic information; see McGettigan & Scott, 2012), with a complementary
5 right-hemisphere dominance for the processing of melodic aspects of spoken signals
6 and the perception of vocal identities (Scott et al., 2000; Kyong et al., 2014;
7 McGettigan & Scott, 2012).

8
9
10
11
12
13
14
15
16
17
18
19 Previous investigations of audiovisual speech intelligibility have shown that the
20 presence of dynamic facial cues improves speech report accuracy under difficult
21 listening conditions (McGettigan, Faulkner, et al., 2012; Ross, Saint-Amour, Leavitt,
22 Javitt, & Foxe, 2007; Sumbly & Pollack, 1954). Neuroimaging studies of audiovisual
23 speech perception have implicated sites including including the posterior superior
24 temporal sulcus (STS), inferior parietal cortex, motor cortex and subcortical structures
25 such as the caudate nucleus (Bernstein, Auer, Wagner, & Ponton, 2008; Calvert et al.,
26 1997; Calvert, Hansen, Iversen, & Brammer, 2001; McGettigan, Faulkner, et al.,
27 2012; Skipper, Nusbaum, & Small, 2005; Stevenson & James, 2009). However,
28 beyond the basic cues to speech intelligibility from the movements of the lips and
29 facial muscles, a talking face brings other information to a communicative interaction,
30 including cues to mood and intentionality - salient amongst these cues is the gaze of
31 the talker. Senju and Johnson (2009) consider the behavioural and neural effects of
32 experiencing eye contact with another person. The authors identify a set of key brain
33 regions that are regularly implicated in studies of gaze perception from faces,
34 including the fusiform gyrus, anterior and posterior portions of the STS, medial
35 prefrontal and orbitofrontal cortices, and the amygdala. They describe how perceived

1 eye contact from another can increase autonomic arousal and modulate activation
2 within the “social brain” (medial prefrontal cortex (mPFC), temporal poles and the
3
4 temporo-parietal junction (TPJ)), thus signaling communicative intent to this system.
5
6 However, they also note inconsistencies in the neuroimaging literature on gaze
7 perception, where some brain regions are only implicated across some studies, while
8
9 other regions show contradictory responses from one study to the next (e.g. mPFC
10 showing a preferential response to direct eye contact in one study, but to averted gaze
11 in another). To make sense of these inconsistencies, Senju and Johnson propose their
12
13 “fast-track modulator” model of eye gaze, in which they suggest that the fundamental
14 mechanism for eye gaze detection is subcortical in its origin, and that the involvement
15
16 of higher-order cortical centres is strongly dependent on task demands.
17
18
19
20
21
22
23
24
25
26
27

28
29 The STS has been a key region implicated in both the perceptual processing of both
30 speech (Binder et al., 2000; Davis & Johnsrude, 2003; Eisner et al., 2010; Evans et
31 al., 2014; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Scott et al., 2000)
32 and eye gaze direction (Calder et al., 2006; Calder et al., 2002; Carlin, Calder,
33
34 Kriegeskorte, Nili, & Rowe, 2011; Hoffman & Haxby, 2000; Hooker et al., 2003;
35
36 Pelphrey, Singerman, Allison, & McCarthy, 2003; Pelphrey, Viola, & McCarthy,
37
38 2004; Puce, Allison, Bentin, Gore, & McCarthy, 1998). Senju and Johnson (2009)
39 describe a difficulty in resolving the relative roles of posterior and anterior sites on
40 the sulcus in terms of the response to eye contact and the mechanisms for
41 discriminating gaze direction, where they suggest that directed attention to the eyes
42 may be required to activate the anterior STS while posterior sites may require
43
44 dynamic visual cues and/or conscious recognition of communicative intent from the
45
46 viewed person. Calder and colleagues carried out a series of studies in which they
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 attempted to better resolve mechanisms for gaze processing along the STS (Calder et
2 al., 2006; Calder et al., 2002; Carlin & Calder, 2013; Carlin et al., 2011). In line with
3 evidence from single-cell recordings in monkeys (e.g. Perrett, Hietanen, Oram, &
4 Benson, 1992; Perrett et al., 1985), Carlin et al. (2011) characterized an anterior-going
5 processing hierarchy in the STS, where posterior sites are sensitive to both gaze and
6 head direction while the anterior STS shows head-direction invariant responses to
7 gaze. In this way, it is suggested that the anterior STS is more responsive to the social
8 significance of gaze than in the specific configurations of the visual cues that signal it.
9 Here, there are strong parallels with the speech perception literature, in which there is
10 a long-standing debate over whether the crucial mapping of sound to linguistic
11 representations takes place primarily in posterior or anterior STS (Evans et al., 2014;
12 Okada et al., 2010; Scott et al., 2000). There is an argument for a speech processing
13 hierarchy in humans that is homologous to the ventral “what” pathway for auditory
14 object recognition in the temporal lobe of non-human primates (Rauschecker & Scott,
15 2009; Scott & Johnsrude, 2003). The anterior STS forms the highest point in this
16 hierarchy, being responsive to intelligible speech signals regardless of their specific
17 acoustic properties (e.g. whether they are undistorted or spectrally degraded; Scott et
18 al., 2000). However, the main distinction between the findings with gaze and speech
19 perception is one of hemispheric lateralization, where responses to speech tend to be
20 left dominant in the superior temporal lobes (McGettigan & Scott, 2012), while
21 sensitivity to gaze direction is more consistently *right-lateralized* (Carlin & Calder,
22 2013).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56 It is tautological to think of “social speech processing”, given that the vast majority of
57 spoken language interactions take place in social settings. Nonetheless, the
58
59
60
61
62
63
64
65

1 neurobiological literature has relatively little to say about social context for spoken
2 communication in terms of how the processing of auditory information might interact
3 with other social cues in speech perception (Scott, McGettigan, & Eisner, 2009).
4 Above, we note the potential commonalities of anterior-going temporal lobe
5 hierarchies for the processing of auditory speech and eye gaze direction. The STS has
6 been repeatedly implicated in the processing of socially-relevant signals, including
7 emotional prosody, facial expressions, vocal identity, gesture and biological motion
8 (Allison, Puce, & McCarthy, 2000; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000;
9 Grandjean et al., 2005; Grezes, Armony, Rowe, & Passingham, 2003). In an attempt
10 to unify this response profile in terms of its underlying computations, Redcay (2008)
11 has proposed that the primary function of the STS is to parse and interpret the
12 communicative significance of incoming streams of audio, visual and audiovisual
13 information unfolding over time. Recent work using vocal signals supports this
14 suggestion of the STS as a locus for social perception, finding that communicative
15 speech and emotional vocalizations generated greater responses in the STS than non-
16 communicative sneezes and coughs (Shultz, Vouloumanos, & Pelphrey, 2012).
17 Similarly, the right posterior STS has been found to be specifically involved in the
18 planning and perception of communicative (vs. non-communicative) actions in a two-
19 player computer game (Noordzij et al., 2010; Stolk et al., 2013).
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 Given the apparent parallels between gaze and speech perception pathways, yet a
49 strong difference in the reported lateralization of these processes, an unanswered
50 question is whether and how the lateralization of one or both processes might be
51 affected by the task context. Behaviourally, there is evidence that heard speech can
52 influence the perception of gaze in a simultaneously presented face – participants
53
54
55
56
57
58
59
60
61
62
63
64
65

1 were more likely to label gaze direction in a static face as direct when an
2 accompanying voice called the participant's name (vs. a control name; Stoyanova,
3 Ewbank, & Calder, 2010). Kampe and colleagues (Kampe, Frith, & Frith, 2003)
4 presented participants with visual and auditory stimuli in separate trials of an event-
5 related fMRI experiment. In the visual condition, they manipulated the gaze of static
6 faces to be direct or averted with respect to the participant, while in the auditory
7 condition a heard voice either demanded the attention of the participant by name (e.g.
8 "Hey John!") or addressed another person. Within each modality, the authors found
9 modulation of key sites in the social brain by conditions with greater communicative
10 intent (i.e. direct gaze and use of the participant's name) - the paracingulate cortex
11 and the left temporal pole were implicated for both modalities. This is an important
12 indication that there is sensitivity to gaze, and its social significance, in the left as well
13 as the right temporal cortex. Using a region of interest analysis, Carlin et al. (2011)
14 also reported head-view invariant responses to gaze direction in the left anterior STS.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 In Senju & Johnson's (2009) proposed "fast-track modulator" model of the eye
37 contact effect, a subcortical stream forms the first path for detection of eye contact
38 and projects to several sites in the social brain. These, in interaction with dorsolateral
39 prefrontal responses to task demands and social context, then influence the further
40 processing of gaze cues in the cortex. [Although this model is focused on the specific
41 percept yielded when another's eyes make direct contact with the gaze of the
42 perceiver, its broad implication is that there is](#) a dynamic network, or set of networks,
43 underpinning the extraction of gaze cues in terms of their social and communicative
44 significance. The aim of the current study was explore the neural responses to talker
45 gaze direction in the context of a speech intelligibility task. We had a particular
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

interest in how the context of a challenging speech task, in which the left hemisphere dominates, would interact with a manipulation that has, in the existing literature, predominately engaged the right temporal lobe. Specifically, we predicted that if participants are primarily engaged in trying to understand what a talker is saying, this could lead to a stronger left-lateralization (or weaker right-lateralization) of superior temporal gaze responses, reflecting dynamic alignment of cortical processes according to the task at hand.

Methods

Materials

Stimuli

The stimuli were 240 English sentences chosen from the BKB (Bamford-Kowal-Bench) list, each featuring three key words (e.g. ‘The CLOWN had a FUNNY FACE’; Bench et al., 1979). The sentences were spoken by a female speaker of Standard Southern British English. The talker, seated, delivered each sentence with three different gaze directions: Direct Gaze (looking directly into the camera lens), Averted Gaze (with gaze held on a single fixed point marked to the right of the camera), Downward Gaze (with head upright, but eyes cast downward toward the talker’s lap).

The videos were filmed in a soundproof room, with the talker’s face set against a blue background and illuminated with a key and a fill light. The talker’s head was fully visible within the frame. Video recordings were made to a Canon XL-1 DV camcorder, with a Bruel & Kjaer type 4165 microphone.

RUNNING HEAD: Neural responses to talker gaze

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

In Final Cut Pro (Apple, Cupertino, CA), the orientation of the raw video was slightly adjusted so that the head was straight and centered. The onset and offset of each sentence was marked (using neutral mouth starting and final positions), and the sentences were exported as individual clips in QuickTime (.mov) format. The mean duration of these raw video clips was 2.99 seconds (s.d. 0.39s, range 1.84-3.96s). In Praat (Boersma & Weenink, 2008), the audio tracks were normalised for peak amplitude, and a cosine ramp was applied at on- and offset. The files were noise-vocoded using a custom-built script in Matlab (The Mathworks, Natick, MA) by passing the speech waveform through a bank of either 2 or 3 analysis filters (upper cutoff 11025Hz). The filter bandwidths were set to represent equal distances along the basilar membrane (according to the Greenwood, 1990 equation relating filter position to best frequency). The amplitude envelope was extracted at the output of each analysis filter via half-wave rectification and low-pass filtering at 400 Hz. The envelopes were each then multiplied with a band-limited white noise carrier, filtered and summed together. The re-summed stimulus was low-pass filtered at 11025 Hz. Finally, the audio versions of each sentence (with 2 or 3 channels) were combined with their respective video using a shell script.

44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

In order to balance the laterality of gaze in the Averted condition, Final Cut Pro was used to horizontally flip half of the sentences in all conditions (Averted, Direct, Downward). Finally, two baseline conditions were created. For each item, a rectangular patch (150 x 150 pixels) was blurred (with radius 150 pixels) using the Joe's Soft Shapes plugin in Final Cut Pro (www.joesfilters.com). The rectangle was positioned horizontally to cover the eyes (Eyes Covered baseline) or vertically to

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

obscure motion cues from the mouth and throat (Mouth Covered baseline). Figure 1 shows an example frame from each of the five visual conditions.

All videos were saved at 720 x 576 pixels in size, at 25 frames per second, with 16-bit audio at sample rate 22050 Hz.

Behavioural Pilot Experiment

Sixteen native speakers of British English (aged 18 – 40 years old) took part in a behavioural sentence report experiment adopting a 5 x 2 factorial design, with the factors Visual Condition (Averted Gaze, Direct Gaze, Downward Gaze, Eyes Covered, Mouth Covered) and Auditory Clarity (2 vs. 3 noise-vocoded channels, where the latter should be of higher intelligibility due to the presence of greater spectral detail). The experiment was conducted with approval from the UCL Research Ethics Committee.

In total, there were 240 trials falling into three blocks of 80 trials. Within each block, there were 10 trials from each of the main experimental conditions (3 gaze directions x 2 levels of auditory clarity), and 5 from each of the mouth and eyes baselines. These were presented in a pseudorandomized order, in which miniblocks of 16 trials featured 2 examples from each of the main experimental conditions and one from each of the baseline conditions (Mouth Covered and Eyes Covered). The stimuli were presented onscreen and over headphones (Sennheiser HD-210, Sennheiser electronic GmbH & Co. KG, Wedermark, Germany) from a MacBook Pro laptop (15" screen, resolution 1440 x 900 pixels; Apple Inc., Cupertino, CA) running Matlab (Version R2009a; The Mathworks, Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997). After each sentence was played, the participant typed what they

1 understood from the sentence, and pressed Enter to advance to the next trial. There
2 was no time limit on responses, and participants were given the opportunity to take
3 breaks between blocks. The participants' responses across all ten conditions were
4 individually scored in terms of the proportion of Key Words that were correctly
5 reported from each sentence.
6
7
8
9
10

11
12
13
14 Figure 2 shows the mean sentence report performance across all conditions). A 5 x 2
15 repeated-measures ANOVA revealed a significant main effect of Auditory Clarity
16 ($F(1,15) = 124.54, p < .001, \text{partial } \eta^2 = .893$) and a significant main effect of
17 Visual Condition ($F(4,60) = 105.20, p < .001, \text{partial } \eta^2 = .875$). Bonferroni-
18 corrected pairwise comparisons revealed that performance on the Mouth Covered
19 condition was significantly worse than on all other visual conditions (all p s $< .001$).
20 The interaction of Auditory Clarity and Visual Condition was non-significant ($F(4,60)$
21 $= 1.34, p = .082, \text{partial } \eta^2 = .082$). A 3 x 2 repeated-measures ANOVA on the six
22 experimental conditions for the MRI experiment with factors Gaze Direction (Direct,
23 Averted, Downward) and Auditory Clarity (Channels: 2 and 3) revealed a significant
24 main effect of Auditory Clarity ($F(1,15) = 104.31, p < .001, \text{partial } \eta^2 = .874$).
25 There was a non-significant effect of Gaze Direction ($F(2,30) = 1.37, p = .270, \text{partial}$
26 $\eta^2 = .084$) and a non-significant interaction of the two factors ($F(2,30) = 1.63, p =$
27 $.214, \text{partial } \eta^2 = .098$).
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 Thus, in line with the existing literature, the behavioural experiment showed that
52 viewing mouth movements can enhance intelligibility of speech in the presence of a
53 degraded auditory signal. However, the direction of gaze had no significant effect on
54 speech comprehension performance. It might have been expected that direct gaze,
55
56
57
58
59
60
61
62
63
64
65

1 being socially salient, might somehow orient attention and thus yield improved
2 sentence report scores – the absence of this effect means, however, that any observed
3 effect of gaze direction in the BOLD signal in STS could not simply be ascribed to
4 differences in linguistic comprehension as a consequence of modulations in visual
5 attention.
6
7
8
9
10

11
12
13
14 *Functional magnetic resonance imaging (fMRI)*
15

16 In the interests of maximizing power in the functional imaging experiment, the Mouth
17 Covered baseline condition used in the behavioural pilot was not included. Thus,
18 participants were presented with a 2 x 3 array of audiovisual sentence conditions with
19 two levels of auditory clarity (2 and 3 channels) and three gaze directions (Direct,
20 Averted, Downward), plus the Eyes Covered baseline (at both levels of auditory
21 clarity) and a rest baseline.
22
23
24
25
26
27
28
29
30

31
32
33
34 *Participants*
35

36 Participants in the study were 18 adults (7 male; mean age 23 years*, range 20-27
37 years) who spoke English as their first language. All were right-handed, with healthy
38 hearing and no history of neurological incidents, nor any problems with speech or
39 language (self-reported). The study was approved by the UCL Research Ethics
40 Committee.
41
42
43
44
45
46
47

48 *Age information was incorrectly recorded for one participant, who does not
49 contribute to this mean. However, information from the participant database confirms
50 that the participant's age was in the range 18-31 years old.
51
52
53
54
55
56
57

58 *Procedure*
59
60
61
62
63
64
65

1 The experiment followed a 2 x 4 design with factors Gaze Direction (Averted,
2 Downward, Direct, None) and Channels (2, 3). A sparse-sampling routine (Edmister,
3 Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999) was employed, in which the
4 audiovisual stimuli were presented in the quiet period between scans. In total, there
5 were 30 trials from each of these eight conditions, organized into miniblocks of 16
6 randomized trials containing 2 examples from each condition. After each miniblock
7 of 16 trials, a seventeenth “catch” trial contained a 1-back vigilance task, in which the
8 participant was probed with an onscreen text keyword and asked to indicate (via
9 keypress on the in-scanner button box) whether this word was contained in the most
10 recently played sentence. The selected keyword came from the preceding trial on 50%
11 of occasions, and on the other 50% was selected randomly from one of the BKB
12 sentences used in the experiment. The catch trial responses were not analysed due to
13 the 50% chance rate and small number of trials (15). In addition to the video and
14 catch trials, there were three miniblocks of silent baseline trials, each lasting for 5
15 trials, which occurred around the midpoint of each functional run. During these trials,
16 the participant saw the words “Mini Break...” written on the screen.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41 Functional imaging data were acquired on a Siemens Avanto 1.5-Tesla MRI scanner
42 (Siemens AG, Erlangen, Germany) with a 32-channel head coil. Audio-visual
43 presentation of sentences took place in three runs of 95 whole-brain volumes using a
44 dual-echo echo-planar imaging sequence (TR = 9 seconds, TA = 3.7 sec, TE = 24; 58
45 ms, 3 mm × 3 mm × 3 mm in-plane resolution, 40 slices with 25 degree tilt transverse
46 to coronal, ascending sequential acquisition).

1 Each video trial began with a 3-second presentation of a fixation cross against a black
2 background. This was positioned roughly at the midpoint between the talker's eyes in
3 the upcoming videos and was presented simultaneously with the onset of the whole-
4 brain volume. Video onsets were timed such that the mid-point of each video
5 occurred 5 seconds before the mid-point of the following whole-brain EPI volume
6 acquisition. By using the variability in durations as a "natural jitter", this resulted in
7 onsets varying across a window of 1.06s (i.e. the difference in the onsets of the
8 longest and shortest videos: $(3.96 - 1.84)/2$).

9
10
11
12
13
14
15
16
17
18
19
20
21 Stimuli were presented using MATLAB with the Psychophysics Toolbox extension,
22 on a MacBook Pro laptop computer (15" screen, resolution 1440 x 900 pixels; Apple
23 Inc., Cupertino, CA). The audio channel was routed through a Sony HD-510 amplifier
24 (Sony Europe Limited, Weybridge, UK) to electrodynamic MR-compatible
25 headphones worn by the participant (Sensimetrics Corporation, Malden, MA).
26
27 Videos were presented at a resolution of 1024 x 768 pixels from an EPSON EH-
28 TW5900 projector (Seiko Epson Corp., Nagano, Japan) to a custom-built screen at the
29 back of the scanner bore, which was viewed using a mirror placed on the head coil.
30
31 Responses to the catch trials were collected via an MR-compatible optical
32 LUMItouch response keypad (Photon Control, Inc., Burnaby, Canada). After the
33 functional run, a high-resolution T1-weighted anatomical image was acquired (HIRes
34 MP-RAGE, 160 sagittal slices, voxel size = 1 mm³).

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53 The total time in the scanner was around 1 hour. As part of the experiment, participant
54 pupil size and gaze direction were measured during the fMRI data acquisition using
55 an Eyelink 1000 MR-compatible eye tracking system (SR Research Ltd., Ottawa,
56
57
58
59
60
61
62
63
64
65

Canada). However, due to technical issues with the labeling of stimulus events, we were unfortunately unable to analyse these data.

Analysis of fMRI data

Analysis of the MRI data was carried out using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). The analysis of EPI data used whole-brain volumes collected on the second echo (TE = 58 ms) only. Functional images were realigned and unwarped, co-registered with the anatomical image, normalized using parameters obtained from unified segmentation of the anatomical image (involving resampling to isometric voxels of 2x2x2mm), and smoothed using a Gaussian kernel of 8 mm FWHM.

At the single-subject level, event onsets from all conditions (4 gaze x 2 channels, plus catch trials) were modeled using a finite impulse response basis function (length: 1 scan, order: 1) in SPM8, along with six movement parameters of no interest. Contrast images for each condition against the implicit baseline (comprising all silent rest trials) were calculated in the single subject and taken forward to a second-level, random effects 3 x 2 within-subjects flexible factorial ANOVA model in SPM8, with factors Subject, Gaze Direction (Direct, Downward, Averted) and Auditory Clarity (2,3). Here, we decided not to model the Eyes Covered conditions in the ANOVA because we did not want to conflate a manipulation of eyes present vs. eyes absent with one of gaze direction. From this 3 x 2 model, F contrast images were calculated for the Main Effect of Gaze ($[\text{kron}([1 \ 1], \text{orth}(\text{diff}(\text{eye}(3))))]$), Main Effect of Auditory Clarity ($[-1 \ -1 \ -1 \ 1 \ 1 \ 1]$) and the Interaction of Gaze Direction and Auditory Clarity ($[\text{kron}([1 \ -1], \text{orth}(\text{diff}(\text{eye}(3))))]$), as well as T contrasts describing the effect of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

increasing Auditory Clarity (3 Channels > 2 Channels; [-1 -1 -1 1 1 1]), the response to Direct Gaze (> Downward and Averted; [2 -1 -1 2 -1 -1]), Averted Gaze (> Direct and Downward; [-1 2 -1 -1 2 -1]) and Downward Gaze (> Averted and Direct; [-1 -1 2 -1 -1 2]), and the combined response to Direct Gaze and Averted Gaze (> Downward; [1 1 -2 1 1 -2]). To allow for an exploration of changes in laterality with speech intelligibility (see below), additional one-way within-subjects ANOVAs with the single factor Gaze Direction (Direct, Downward, Averted) were run separately for the two levels of Auditory Clarity (2 Channels, 3 Channels). Finally, to allow for pairwise comparisons of gaze conditions and their interactions with speech intelligibility, three within-subjects ANOVAs with factors Gaze Direction and Auditory were run for (i) Direct vs. Downward, (ii) Averted vs. Downward and (iii) Direct vs. Averted – the results of these analyses can be found in the Supplemental Material.

All second-level models were calculated at a voxelwise threshold of $p < .005$ (uncorrected). A cluster extent correction of 68 voxels (544mm^3) was applied for a whole-brain alpha of $p < .001$ using a Monte Carlo simulation (with 10 000 iterations) implemented in MATLAB (with smoothness estimate of 13.2mm; Slotnick, Moo, Segal, & Hart, 2003).

Second-level peak coordinates were used to extract condition-specific parameter estimates from 4mm-radius spherical regions of interest (ROIs) built around the peak voxel (using MarsBaR; Brett, Anton, Valabregue, & Poline, 2002). The anatomical locations of peak and sub-peak voxels (at least 8mm apart) were labelled using the SPM Anatomy Toolbox (version 20) (Eickhoff et al., 2005).

Calculating Laterality Indices

1
2 To test the temporal lobe lateralization of activation in the Main effect of Gaze
3
4 Direction, Main Effect of Auditory Clarity, Positive Effect of Auditory Clarity ($2 < 3$)
5
6 and the preferential response to Direct gaze ($>$ Averted and Downward), we used the
7
8 LI toolbox in SPM8 ([https://www.medizin.uni-](https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/?download=li-toolbox)
9
10 [tuebingen.de/kinder/en/research/neuroimaging/software/?download=li-toolbox](https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/?download=li-toolbox); Wilke
11
12 & Schmithorst, 2006). For each contrast of interest, the toolbox calculates laterality
13
14 indices (LI) using the equation: $LI = (\Sigma_{activationleft} - \Sigma_{activationright} /$
15
16 $\Sigma_{activationleft} + \Sigma_{activationright})$, where Σ refers to the sum of activation either in
17
18 terms of the total voxel count, or the sum of the voxel values within the statistical map
19
20 of the contrast. Thus, values, of LI can vary from +1 (completely left lateralized) to -1
21
22 (completely right lateralized). According to convention, an absolute LI value greater
23
24 than 0.2 is taken to indicate a hemispheric dominance (Seghier, 2008). In this paper,
25
26 “activation” in the LI formula was defined as the total voxel values within each
27
28 hemisphere in the second-level 3 x 2 flexible factorial ANOVA F maps of the Main
29
30 Effect of Gaze Direction and Main Effect of Auditory Clarity, and the T maps of the
31
32 Positive Effect of Auditory Clarity and preferential response to Direct gaze (over
33
34 Averted and Downward), restricted in our case to the left and right temporal lobes
35
36 (defined using an inclusive bilateral anatomical mask of the superior and middle
37
38 temporal gyri and temporal poles constructed from the AAL regions of interest
39
40 available in the Marsbar toolbox; Brett et al., 2002). LIs were also calculated for the
41
42 contrast Direct Gaze $>$ (Averted and Downward) for separate one-way within subject
43
44 ANOVAs using conditions with 2 channels only, and 3 channels only. To take
45
46 account of thresholding effects, the toolbox calculates LIs at 20 thresholding intervals
47
48 from 0 to the maximum value in the F/T map. At each level, the toolbox selects 100
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 bootstrap samples (5-1000 voxels) from each masked hemisphere, which are paired in
2 all possible combinations (10,000) and used to calculate an equivalent number of LIs.
3
4 From the final distribution of LIs, the toolbox reports trimmed means (where the top
5 and bottom 25% of values have been discarded), as well as a single weighted mean
6 based on these that is proportionally more affected by LI values from higher statistical
7 thresholds. Here, we report the trimmed and weighted means for each contrast of
8 interest.
9
10
11
12
13
14
15
16
17
18

19 **Results**

20
21 *The left temporal lobe is preferentially responsive to increasing intelligibility of*
22 *degraded speech*
23
24
25

26 A contrast exploring the main effect of Auditory Clarity revealed significant clusters
27 in regions of superior temporal sulcus (both posterior and anterior) and inferior frontal
28 (extending into premotor) cortex, as found in several previous studies (Davis &
29 Johnsrude, 2003; Eisner et al., 2010; McGettigan, Evans et al., 2012; McGettigan et
30 al., 2012). (see Figure 3 and Table 1). These all showed a similar profile, where there
31 were larger responses to sentences with 3 noise-vocoded channels (and hence greater
32 auditory clarity) than to sentences with 2 noise-vocoded channels. The activation in
33 the superior temporal lobes was strongly left lateralized (according to the weighted
34 conventional threshold of >0.2 ; Seghier et al., 2008. See Table 2).
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 Additional activations for this effect of Auditory Clarity (and for the directional T-
52 contrast of 3 Channels $>$ 2 Channels; see Table 1) included superior frontal cortex, a
53 large cluster in bilateral calcarine gyrus and cuneus, bilateral sensorimotor cortex and
54 regions of medial prefrontal cortex including the cingulate gyrus.
55
56
57
58
59
60
61
62
63
64
65

1
2 *The left anterior temporal lobe is preferentially sensitive to socially salient gaze*
3
4 *during speech perception*
5
6

7 Figure 4 illustrates regions of significant activation for the main effect of Gaze
8 Direction (see also Table 1), including regions of visual, parietal and prefrontal
9 cortex. The general trend was for greater responses to conditions where the sclera and
10 pupils were visible; that is, the Direct and Averted Gaze conditions. Several regions
11 were found to be more active in response to sentences in which the talker was looking
12 directly at the camera (and, hence, the viewer) compared with averted and downward
13 gaze (Figure 5). Many of these regions overlapped with significant clusters in a
14 contrast exploring shared preference for either direct or averted gaze (Table 1, Figure
15 6), and included areas of inferior and superior frontal, visual, sensorimotor and
16 parietal cortex. However, a region in the anterior temporal lobe (extending to
17 temporal pole) showed a distinct preference for direct gaze and, like the preferential
18 response to increased auditory clarity, showed a strong left lateralization (Table 2). In
19 the 3 x 2 ANOVA model, All weighted LIs for the Main Effect of Gaze Direction,
20 Main Effect of Auditory Clarity, 3 Channels > 2 Channels and Direct > (Averted and
21 Downward) contrasts were strongly left dominant in the superior temporal lobe (See
22 Table 2). Preferential responses to averted gaze direction were mainly confined to
23 regions of the right superior frontal and parietal cortices (Table 1).
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 There were no significant clusters showing an interaction of Auditory Clarity and
52 Gaze Direction in the 3 x 2 ANOVA. However, analysis of temporal lobe
53 lateralization within the T maps for the contrast Direct > (Averted and Downward) at
54 the two different levels of Auditory Clarity showed that this effect was strongly left
55
56
57
58
59
60
61
62
63
64
65

1 lateralized at 2 Channels (weighted mean = 0.44) but showed no lateralization at 3
2 Channels (weighted mean = -0.09; see Tables 2 and 3). This suggests that when there
3 were fewer available cues to the acoustic content of speech, participants' responses to
4 direct gaze were more strongly expressed in the left temporal lobe (presumably due to
5 greater attention to the face to assist the speech comprehension task).
6
7
8
9
10

11
12
13
14 *Bilateral anterior insula and inferior frontal cortex show common responses to*
15 *increases in intelligibility and perceived communicative intent of audio-visual speech.*

16
17 Figure 5 overlays the significant activations showing preferential responses to direct
18 gaze and increased auditory clarity, where overlap can be seen in anterior temporal
19 cortex, bilateral calcarine gyrus, and bilateral inferior frontal cortex and insula. A
20 conjunction null of these two contrasts resulted in four significant clusters, with peaks
21 in left inferior frontal gyrus, right insula, left lingual gyrus and right calcarine gyrus.
22
23
24
25
26
27
28
29
30
31

32 **Discussion**

33
34 The main aim of the study was to explore whether task context could influence the
35 hemispheric lateralization of neural responses to the perception of the talker's eye
36 gaze in the temporal lobe, specifically the STS. Indeed, we observed a main effect of
37 talker gaze direction (specifically, a preferential response to direct gaze in the anterior
38 temporal lobe) that was strongly left-lateralized and partially overlapping with a
39 similar anterior temporal response to increases in auditory intelligibility. On closer
40 investigation, we found that the preferential response to direct gaze in anterior
41 temporal cortex was strongly left-lateralized at lower levels of auditory speech
42 intelligibility, but showed no hemispheric dominance when auditory clarity was
43 increased. This suggests flexibility in the spatial distribution of processes across
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 homologous brain regions (i.e. gaze processing in the right and left anterior temporal
 2 lobes; Carlin et al., 2011) during spoken communication, where in this case the
 3 responses to speech and [direct gaze](#) were expressed most strongly in the same cerebral
 4 hemisphere when attention to both the face and the voice were maximally important
 5 for performance of the task. Thus, in line with Senju & Johnson's (2009) proposal¹,
 6 we present the evidence for dynamic, task-dependent responses to gaze at this
 7 relatively early stage in the speech comprehension process (i.e. the extraction of
 8 speech from a degraded auditory input). However, this conclusion should be tempered
 9 somewhat by the lack of significant interaction in temporal cortex between the gaze
 10 and intelligibility manipulations in our main ANOVA on the fMRI data.

26 Observing a response to gaze direction in the left anterior temporal lobe is not
 27 completely at odds with previous research, despite that fact that this has emphasized
 28 right-lateralized effects (Calder et al., 2002; Pelphrey et al., 2004; Calder et al., 2006;
 29 Carlin et al., 2011; Carlin & Calder, 2013). Carlin and colleagues found head view-
 30 invariant responses to gaze direction in the right anterior STS, but using a region-of-
 31 interest analysis found that this response was also present in the homologous part of
 32 the left hemisphere. Kampe et al. (2003) found that the conjunction of activations in
 33 response to auditory and visual communicative signals revealed a cluster in the left
 34 temporal pole (and not the right). Our results fall in line with previous work
 35 suggesting that the anterior STS is the locus of extracting the social meaning of gaze
 36 (Carlin & Calder, 2013) – however, based on the work of Carlin and colleagues, we
 37 did not predict engagement of posterior sites because our talker maintained the same,

56 ¹ Here, we refer only to Senju & Johnson's (2009) descriptions of task-dependent
 57 responses to gaze direction rather than to eye contact - due to a lack of usable eye
 58 tracking data, we cannot tease apart direct gaze perception from the perception eye
 59 contact in the current study.

1 front-facing head position throughout. Other authors have previously discussed the
 2 STS as a site for the perception of social and communicative signals in the voice
 3 (Redcay, 2008; Shultz et al., 2012), and the left-lateralization of gaze and speech
 4 responses identified in this study suggests a potential common processing pathway in
 5 this region for extracting meaning from another’s actions (here, the movements of the
 6 mouth and the eyes during spoken communication). Notably, the BOLD signal in
 7 response to Downward gaze tended to be the smallest of the three gaze conditions
 8 (see plots of parameter estimates in Figures 3 and 4), and indeed lower than the Eyes
 9 Covered baseline. It could be argued that a downward glance was the least
 10 communicative visual stimulus presented - in the baseline stimuli, the participants
 11 may have maintained some expectation that a communicative gaze was being
 12 obscured by the masking rectangle (a reason why we did not include the Eyes
 13 Covered conditions in the Gaze Direction factor of our ANOVA analyses).

14 Direct gaze also engaged other structures previously implicated in studies employing
 15 gaze manipulations, including calcarine and fusiform gyri, and bilateral insula/IFG
 16 (Calder et al., 2002; Callejas, Shulman, & Corbetta, 2014; Ethofer, Gschwind, &
 17 Vuilleumier, 2011; Pitskel et al., 2011). We did not observe ventromedial prefrontal
 18 cortex, as seen by Kampe et al. (2003) and Calder et al. (2002). Kuzmanovic and
 19 colleagues (2009) found greater activation of medial prefrontal cortex in response to
 20 direct gaze of increasing duration, and Bristow and colleagues (2007) found that
 21 regions of the “social brain” in medial prefrontal cortex and precuneus were engaged
 22 by attentional cueing gaze shifts that followed socially salient direct eye contact. The
 23 lack of medial prefrontal engagement in the current study may be an effect of the task
 24 demands, where our participants’ primary goal was to extract and comprehend the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

linguistic message and not to make an overt social judgement. Even in the absence of a social task, many previous studies have presented visual-only stimuli, and so the modulation of visual social cues was more prominent in those studies. Furthermore, the BKB sentences used in our study, which either refer to people in the 3rd person (e.g. “They’re pushing an old car”) or to scenes without human agents (e.g. “The rain came down”) did not provide a strong sense of social context.

Averted gaze can have strong social and informational salience. A glance to the side can signal that another individual is being addressed (e.g. Holler et al., 2015), or that the talker wishes the viewer to pay attention to a particular part of space (e.g. Bristow et al., 2007; Pelphrey et al., 2003; Ramsey, Cross, & Hamilton, 2011). It can also potentially signal internal states in the talker (e.g. direct gaze in anger, averted gaze in sadness; Adams & Kleck, 2003). In the current experiment, preferential responses to averted gaze were found in right-dominant regions of the superior parietal lobe (including the intraparietal sulcus) and bilateral superior frontal gyri. This may correspond to engagement of the dorsal attentional network (DAN; Callejas et al., 2014; Corbetta, Patel, & Shulman, 2008), where averted gaze stimulates the expectation of an object or event at a lateralized location in space (Ramsey et al., 2011). In a simple sentence comprehension task, Staudte and colleagues (2014) observed that (averted) gaze direction aids comprehension through visual cueing (e.g. signaling the location of an object) and not by signaling the speaker’s intentions. As the current study provided no explicit social context associated with averted gaze, our averted stimuli may similarly have been perceived as attentionally directing rather than strongly communicative (compared with direct gaze stimuli).

1 Our study revealed sensitivity to both the direction of gaze and the availability of
2 additional auditory information in parts of the anterior insula bilaterally, extending
3 into the inferior frontal gyrus. Similarly to effects in the temporal lobe, the
4 involvement of the anterior insula and inferior frontal cortex in the perceptual
5 processing of speech has tended to be left dominant (Davis & Johnsrude, 2003; Eisner
6 et al., 2010; McGettigan, Evans, et al., 2012). Engagement of (bilateral) inferior
7 frontal sites in the current experiment could partly reflect an overall stronger focus on
8 the linguistic task in speech-processing regions in response to attentional capture by
9 the eyes. The perception of direct eye contact is highly salient (Senju & Johnson,
10 2009) and communicatively important (Kampe et al., 2003) - the ventral attention
11 network (VAN), implicating inferior frontal cortex, has been shown to be responsive
12 to the occurrence of unexpected and important events (Corbetta et al., 2008).
13 Inclusion of an attentional task demands, in future work, for example to attend to the
14 mouth on some trials and the eyes on others, may allow us to tease apart cue-specific
15 from domain-general aspects of attentional engagement with gaze in our stimuli. [Such
16 studies will be dependent on eye tracking data to verify adherence to the task \(and to
17 allow us to test predictions about eye contact vs. general gaze perception; see
18 Footnote 1\).](#)

19 In recent years, an emerging literature has employed manipulations of gesture, gaze
20 and body posture to investigate language comprehension in different communicative
21 contexts. As in our study, there is evidence for varied distribution of activation
22 depending on task and stimulus contexts. An enhanced response to direct gaze during
23 spoken sentence comprehension was observed in right MTG (Holler et al., 2015), but
24 only for conditions including speech and gesture (and not speech alone) – this was

1 interpreted as evidence for the integration of meaning from multimodal inputs in the
2 right temporal lobe. Straube and colleagues (2010) showed widespread enhancement
3 of BOLD responses to stimuli in which the participant was directly addressed (i.e.
4 speaker facing toward the viewer), which was larger in bilateral anterior temporal
5 regions for sentences containing person-related (versus object-related) information.
6 That these studies did not show similar left-dominant temporal responses to direct
7 gaze, as seen in our study, may be due to a number of factors - the increased difficulty
8 of our speech comprehension task potentially loaded more strongly on lower-level
9 speech perception processes, and the use of face-only (rather than whole body) visual
10 stimuli may have resulted in more focused attention on gaze direction. Further, it
11 could be argued that previous studies using unmanipulated (i.e. clear) speech, as well
12 as contrasts of semantic and pragmatic sentence properties, might measure relatively
13 higher-order social and cognitive processes than the current study. However, it is also
14 important to note here that while we have identified evidence that left-lateralization of
15 responses to gaze direction became stronger when auditory intelligibility was reduced,
16 thus suggesting an effect of speech processing on the response to gaze, we cannot
17 make claims about the overall left-lateralization *per se* in relation to the speech
18 communicative context. In order to support such a claim, we would have had to
19 include some visual-only stimuli, and a non-speech baseline task – such modifications
20 of the design should be implemented in future work.

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51 Several recent studies have more closely examined social processing in language
52 comprehension by considering the effects of participants' subjective experience of the
53 contextual manipulations. For example, Nagels and colleagues (2015) found that
54 activations in the anterior cingulate cortex, fusiform gyrus, SMA, IFG and insula were
55
56
57
58
59
60
61
62
63
64
65

1 positively correlated with subjective feelings of being addressed by the speaker
2 (which was varied through manipulations to the talker's body posture and the use of
3 co-speech gestures). In a study of gesture and written language comprehension, it was
4 found that responses to participant-directed gestures (compared with non-
5 communicative gestures) in regions of left STS were related to participants'
6 subjective ratings of communicative intent in the gesture stimuli – the same regions
7 also showed greater responses to participant-directed sentences (compared with 3rd-
8 person sentences; Redcay, Velnoskey & Rowe, 2016). Finally, a recent study that
9 modulated the participant's belief about the presence of live and recorded
10 interlocutors showed increased signal within mentalizing regions including mPFC and
11 TPJ, and that the strength of the participant's subjective experience of the "live"
12 speech was associated with increased responses across the mentalizing network (Rice
13 & Redcay, 2016). Although it is accepted that direct gaze forms a salient social and
14 communicative cue, the current study did not present participants with an explicit
15 social context for the different gaze conditions used in the task. Future developments
16 of this work could involve more direct manipulation of context (e.g. presenting the
17 Averted Gaze condition as signaling speaker intention to another viewer; Holler et al.,
18 2015) as well as manipulation of the sentence content to compare higher and lower
19 levels of participant-directedness. Crucially, collecting subjective ratings of
20 communicative intent would allow for stronger claims as to the social significance of
21 gaze and sentence content, which could potentially be explored at a wider range of
22 speech intelligibility levels than explored in the current experiment.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Conclusions

1 We report the first neuroimaging study to explore the interaction of talker gaze
2 direction with the extraction of intelligible speech from degraded auditory inputs. We
3
4 found that, when listening to degraded speech, the neural response to direct gaze in
5
6 the STS that has previously been reported as predominately right-lateralized was
7
8 lateralized to the left hemisphere, specifically at lower levels of auditory speech
9
10 intelligibility. This finding supports Senju & Johnson's (2009) view of flexibility in
11
12 the distribution of gaze-related activations around the cerebral cortex, depending on
13
14 task demands, and with the argument that the human STS is optimized for the
15
16 processing of social cues (Redcay, 2008). While the literature to date has associated
17
18 speech more closely with linguistic computations, we argue that speech is
19
20 fundamentally social, and that future research should aim to more closely examine
21
22 how the context of spoken interactions both affords and affects social processing
23
24 across the brain. More widely, our view aligns with an emerging movement in the
25
26 social neuroscience literature arguing that, in order to truly understand the social
27
28 brain, we should strive to investigate social processes in real time and within
29
30 naturalistic interactions (a "second-person neuroscience": e.g. Schilbach et al., 2013;
31
32 McGettigan, 2015).
33
34
35
36
37
38
39
40
41
42

43 **Funding**

44 This work was supported by a Wellcome Trust Senior Research Fellowship
45
46 (WT090961MA) awarded to Sophie Scott.
47
48
49
50

51 **References**

RUNNING HEAD: Neural responses to talker gaze

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, 14(6), 644-647. doi: Doi

10.1046/J.0956-7976.2003.Psci_1479.X

Adank, P. (2012a). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42-54. doi: 10.1016/j.bandl.2012.04.014

Adank, P. (2012b). Design choices in imaging speech comprehension: an Activation Likelihood Estimation (ALE) meta-analysis. *NeuroImage*, 63 (3), 1601-1613. doi:10.1016/j.neuroimage.2012.07.027

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267-278. doi: Doi 10.1016/S1364-6613(00)01501-1

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309-312. doi: 10.1038/35002078

Bernstein, L. E., Auer, E. T., Wagner, M., & Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing. *Neuroimage*, 39(1), 423-435. doi: Doi 10.1016/J.Neuroimage.2007.08.035

Bench, J., Kowal, A., Bamford, J. (1979) The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology* 13(3):108-12. DOI: 10.3109/03005367909078884

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*, 10(5), 512-528.

Boersma, Paul & Weenink, David (2008). Praat: doing phonetics by computer [Computer program]. Retrieved 1 January 2008 from <http://www.praat.org/>

RUNNING HEAD: Neural responses to talker gaze

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436.

Brett M, Anton JL, Valabregue R, Poline JB. Region of interest analysis using an SPM toolbox. *Neuroimage*. 2002;16(2):1140–1141.

Bristow, D., Rees, G., & Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Social Cognitive and Affective Neuroscience*, 2(1), 52-61.

doi: Doi 10.1093/Scan/Nsl036

Calder, A. J., Beaver, J. D., Eger, E., Jenkins, R., Winston, J., Dolan, R. J., & Henson, R. N. A. (2006). The neural correlates of eye gaze adaptation. *Perception*, 35, 240-240.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, 40(8), 1129-1138. doi: Pii S0028-3932(02)00008-8

Doi 10.1016/S0028-3932(02)00008-8

Callejas, A., Shulman, G. L., & Corbetta, M. (2014). Dorsal and Ventral Attention Systems Underlie Social and Symbolic Cueing. *Journal of Cognitive Neuroscience*, 26(1), 63-80. doi: Doi 10.1162/Jocn_a_00461

Calvert, G., Woodruff, P., Wright, I., Bullmore, E., Brammer, M., Williams, S., . . .

David, A. (1997). fMRI imaging of visual and audiovisual speech. *International Journal of Psychophysiology*, 25(1), 23-23. doi: Doi 10.1016/S0167-8760(97)85392-5

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, 14(2), 427-438. doi: Doi 10.1006/Nimg.2001.0812

Carlin, J. D., & Calder, A. J. (2013). The neural basis of eye gaze processing. *Current Opinion in Neurobiology*, 23(3), 450-455. doi: Doi 10.1016/J.Conb.2012.11.014

RUNNING HEAD: Neural responses to talker gaze

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A Head View-Invariant Representation of Gaze Direction in Anterior Superior Temporal Sulcus. *Current Biology*, 21(21), 1817-1821. doi: Doi 10.1016/J.Cub.2011.09.025

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, 58(3), 306-324. doi: Doi 10.1016/J.Neuron.2008.04.017

Davis, M. H., & Johnsrude, I. S. (2003a). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423-3431.

Davis, M. H., & Johnsrude, I. S. (2003b). Hierarchical processing in spoken language comprehension. *J Neurosci*, 23(8), 3423-3431.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999).

Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp*, 7(2), 89-97. doi: 10.1002/(sici)1097-0193(1999)7:2<89::aid-hbm2>3.0.co;2-n

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations. *Journal of Neuroscience*, 30(21), 7179-7186. doi:

10.1523/jneurosci.4040-09.2010

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic

cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325-1335.

Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: A combined fMRI-DTI study. *Neuroimage*, 55(1), 411-419. doi: Doi 10.1016/J.Neuroimage.2010.11.033

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., . . .

Scott, S. K. (2014). The pathways for intelligible speech: multivariate and univariate

perspectives. [Research Support, Non-U.S. Gov't]. *Cereb Cortex*, 24(9), 2350-2361.

doi: 10.1093/cercor/bht083

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 8(2), 145-146. doi: 10.1038/nn1392

Greenwood, D.D. (1990) A cochlear frequency-position function for several species--29 years later. *Journal of the Acoustical Society of America*. 87(6):2592-605.

Grezes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study.

Neuroimage, 18(4), 928-937. doi: Doi 10.1016/S1053-8119(03)00042-9

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q.,

Elliott, M. R., . . . Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory

fMRI. *Hum Brain Mapp*, 7(3), 213-223. doi: 10.1002/(sici)1097-

0193(1999)7:3<213::aid-hbm5>3.0.co;2-n

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012).

Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang*

Cogn Process, 27(7-8), 1145-1166. doi: 10.1080/01690965.2012.662280

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing.

Nat Rev Neurosci, 8(5), 393-402.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and

identity in the distributed human neural system for face perception. *Nature*

Neuroscience, 3(1), 80-84. doi: Doi 10.1038/71152

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S.D., Ozyurek, A. (2015). Eye'm

talking to you: speakers' gaze direction modulates co-speech gesture processing in the

RUNNING HEAD: Neural responses to talker gaze

right MTG. *Social Cognitive and Affective Neuroscience* 10(2), 255-61. doi:

10.1093/scan/nsu047

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, 17(2), 406-418. doi: Doi 10.1016/S0926-6410(03)00143-5

Kampe, K. K. W., Frith, C. D., & Frith, U. (2003). "Hey John": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *Journal of Neuroscience*, 23(12), 5258-5263.

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., & Vogeley, K. (2009). Duration matters: Dissociating neural correlates of detection and evaluation of social gaze. *Neuroimage*, 46(4), 1154-1163. doi: Doi 10.1016/J.Neuroimage.2009.03.037

Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the Roles of Spectral Detail and Intonation Contour in Speech Intelligibility: An fMRI Study. *Journal of Cognitive Neuroscience*, 26(8), 1748-1763. doi: Doi 10.1162/Jocn_a_00583

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb Cortex*, 15(10), 1621-1631. doi: bhi040 10.1093/cercor/bhi040

McGettigan C (2015) The social life of voices: Studying the neural bases for the expression and perception of the self and others during spoken communication. *Frontiers in Human Neuroscience* 9, 129.

McGettigan, C., Evans, S., Rosen, S., Agnew, Z. K., Shah, P., & Scott, S. K. (2012). An application of univariate and multivariate approaches in FMRI to quantifying the

RUNNING HEAD: Neural responses to talker gaze

hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience*, 24(3), 636-652. doi: 10.1162/jocn_a_00161

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, 50(5), 762-776. doi: 10.1016/j.neuropsychologia.2012.01.010

McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends in Cognitive Sciences*, 16(5), 269-276. doi: 10.1016/j.tics.2012.04.006

Nagels, A., Kircher, T., Steines, M., & Straube, B. (2015). Feeling addressed! The role of body orientation and co-speech gesture in social communication. *Human Brain Mapping*, 36(5), 1925–36. doi:10.1002/hbm.22746

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex*, 13(12), 1362-1368.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., Toni, I. (2010). Neural correlates of intentional communication. *Frontiers in Neuroscience* 4, 188.

Obleser, J., Wise, R. J., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci*, 27(9), 2283-2289.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., . . . Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex*, 20(10), 2486-2495. doi: 10.1093/cercor/bhp318

1
2 Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain
3 activation evoked by perception of gaze shifts: the influence of context.

4
5 Neuropsychologia, 41(2), 156-170. doi: Pii S0028-3932(02)00146-X

6
7 Doi 10.1016/S0028-3932(02)00146-X

8
9 Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass -
10 Processing of mutual and averted social gaze in the superior temporal sulcus.

11
12 Psychological Science, 15(9), 598-603. doi: Doi 10.1111/J.0956-7976.2004.00726.X

13
14 Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and
15 Functions of Cells Responsive to Faces in the Temporal Cortex. Philosophical

16
17 Transactions of the Royal Society of London Series B-Biological Sciences,

18
19 335(1273), 23-30. doi: Doi 10.1098/Rstb.1992.0003

20
21 Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D.,

22
23 & Jeeves, M. A. (1985). Visual Cells in the Temporal Cortex Sensitive to Face View
24 and Gaze Direction. Proceedings of the Royal Society Series B-Biological Sciences,

25
26 223(1232), 293-317. doi: Doi 10.1098/Rspb.1985.0003

27
28 Pitskel, N. B., Bolling, D. Z., Hudac, C. M., Lantz, S. D., Minshew, N. J., Vander

29
30 Wyk, B. C., & Pelphrey, K. A. (2011). Brain Mechanisms for Processing Direct and

31
32 Averted Gaze in Individuals with Autism. Journal of Autism and Developmental

33
34 Disorders, 41(12), 1686-1693. doi: Doi 10.1007/S10803-011-1197-X

35
36 Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal

37
38 cortex activation in humans viewing eye and mouth movements. Journal of

39
40 Neuroscience, 18(6), 2188-2199.

41
42 Ramsey, R., Cross, E. S., & Hamilton, A. F. D. (2011). Eye Can See What You Want:

43
44 Posterior Intraparietal Sulcus Encodes the Object of an Actor's Gaze. Journal of

45
46 Cognitive Neuroscience, 23(11), 3400-3409.

RUNNING HEAD: Neural responses to talker gaze

1 Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex:
2 nonhuman primates illuminate human speech processing. *Nat Neurosci*, 12(6), 718-
3 724.
4
5

6 Redcay, E. (2008). The superior temporal sulcus performs a common function for
7 social and speech perception: Implications for the emergence of autism. *Neuroscience*
8 and *Biobehavioral Reviews*, 32(1), 123-142. doi: 10.1016/j.neubiorev.2007.06.004
9
10

11 Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016) Perceived Communicative
12 Intent in Gesture and Language Modulates the Superior Temporal Sulcus. *Human*
13 *Brain Mapping* 37, 3444-3461.
14
15

16 Rice, K., Redcay, E. (2016) Interaction matters: A perceived social partner alters the
17 neural processing of human speech. *NeuroImage* 129, 480-488.
18
19

20 Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do
21 you see what I am saying? Exploring visual enhancement of speech comprehension in
22 noisy environment. *Cerebral Cortex*, 17(5), 1147-1153. doi: Doi
23 10.1093/Cercor/Bhl024
24
25

26 Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T.,
27 Vogeley, K. (2014) Toward a second-person neuroscience. *Behavioral and Brain*
28 *Sciences*. 36(4):393-414. doi: 10.1017/S0140525X12000660.
29
30

31 Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway
32 for intelligible speech in the left temporal lobe. *Brain*, 123, 2400-2406.
33
34

35 Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional
36 organization of speech perception. *Trends Neurosci*, 26(2), 100-107. doi:
37 S0166223602000371
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

RUNNING HEAD: Neural responses to talker gaze

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action--candidate roles for the motor cortex in speech perception. *Nat Rev Neurosci*, 10(4), 295-302.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *J Acoust Soc Am*, 120(2), 1075-1083.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, 13(3), 127-134. doi: Doi 10.1016/J.Tics.2008.11.009

Seghier, M.L. (2008) Laterality index in functional MRI: methodological issues. *Magnetic Resonance Imaging*. 26(5): 594–601. doi: 10.1016/j.mri.2007.10.010

Shultz, S., Vouloumanos, A., & Pelphrey, K. (2012). The Superior Temporal Sulcus Differentiates Communicative and Noncommunicative Auditory Signals. *Journal of Cognitive Neuroscience*, 24(5), 1224-1232.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, 25(1), 76-89. doi: 10.1016/j.neuroimage.2004.11.006

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, 17(1), 75-82. doi: 10.1016/s0926-6410(03)00082-x

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44(3), 1210-1223. doi: Doi 10.1016/J.Neuroimage.2008.09.034

1 Stolk, A., Verhagen, L., Schoffelen, J., Oostenveld, R., Blokpoel, M., Hagoort, P.,
2 van Rooij, I., & Toni, I. (2013) Neural mechanisms of communicative innovation.

3
4
5 Proceedings of the National Academy of Sciences USA, 110, 14574-14579.

6
7 Stoyanova, R. S., Ewbank, M. P., & Calder, A. J. (2010). "You Talkin' to Me?": Self-
8
9 Relevant Auditory Signals Influence Perception of Gaze Direction. Psychological
10
11 Science, 21(12), 1765-1769. doi: Doi 10.1177/0956797610388812

12
13
14 Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues,
15
16 mentalizing and the neural processing of speech accompanied by gestures.

17
18
19 Neuropsychologia, 48(2), 382–393. doi:10.1016/j.neuropsychologia.2009.09.025

20
21
22 Sumbly, W., Pollack, I. . (1954). Visual Contribution to Speech Intelligibility in Noise.
23
24 Journal of the Acoustical Society of America, 26, 212–215.

25
26
27 Wilke M, Schmithorst VJ. (2006) A combined bootstrap/histogram analysis approach
28
29 for computing a lateralization index from neuroimaging data. NeuroImage 33(2), 522-
30
31 30.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Tables

Table 1 Results of the main contrasts exploring the effects of Gaze Direction and Auditory Clarity. All results are reported at a voxel height threshold of $p < .005$ (uncorrected), and a corrected cluster threshold of $p < .001$ (Slotnick et al., 2003). Coordinates are given in Montreal Neurological Institute (MNI) stereotactic space.

Contrast	No of Voxels	Region(s)	Peak Coordinate			F/T	Z
			x	y	z		
Main Effect of Gaze Direction (F-test)	2680	Right/left calcarine gyrus and left cuneus	24	-62	10	16.25	4.74
	69	Right inferior parietal lobule	62	-50	44	14.17	4.42
	186	Right IFG (pars triangularis)	36	28	2	13.95	4.39
	111	Left cerebellum (lobules IV-V)	-12	-50	-12	10.91	3.84

RUNNING HEAD: Neural responses to talker gaze

	119	Left STG/STS/temporal pole	-62	-2	-6	10.87	3.84
	118	Left precuneus	-6	-80	46	10.06	3.67
	316	Right superior and inferior parietal lobule; right post-central gyrus	20	-48	58	9.76	3.61
	68	Right postcentral gyrus	60	-4	20	8.93	3.43
	115	Right superior/middle frontal gyrus	26	-2	54	8.79	3.40
	82	Left posterior-medial frontal lobe; left mid-cingulate cortex	-8	10	46	7.99	3.21
Main Effect of Auditory Clarity (F-test)	1031	Left IFG (pars triangularis, orbitalis); left precentral gyrus; left insula	-50	18	22	19.13	
	223	Left superior/middle frontal gyrus	-22	-6	52	18.18	3.88
	136	Left superior frontal gyrus	-10	52	36	17.13	3.77
	1658	Left STS/STG; left middle occipital gyrus	-40	-52	18	17.00	3.76
	166	Right STS/STG	50	-16	-6	16.48	3.70
	227	Bilateral posterior-medial frontal lobe	4	20	50	16.26	3.67
	260	Left mid-cingulate cortex; right posterior-medial frontal lobe	0	0	34	15.79	3.62
	247	Right IFG (pars triangularis); right insula	44	30	6	15.35	3.57

RUNNING HEAD: Neural responses to talker gaze

	104	Left lingual gyrus	-14	-56	2	15.33	3.57
	195	Left medial temporal lobe (white matter)	-28	-50	8	12.91	3.27
	68	Left postcentral gyrus	-32	-38	44	12.94	3.21
	113	Left central sulcus/post-central gyrus	-24	-28	52	11.99	3.14
	150	Bilateral cuneus	4	-52	62	11.85	3.12
	73	Right precentral gyrus/central sulcus	56	-14	46	11.65	3.10
	240	Left STG/STS	-54	-8	-2	11.57	3.08
3 channels > 2 channels (T- test)	1530	Left IFG (pars triangularis, orbitalis); left precentral gyrus; left insula	-50	18	22	4.37	4.14
	343	Left superior/middle frontal gyrus	-22	-6	52	4.26	4.05
	353	Left superior frontal gyrus; bilateral superior medial gyrus; right middle frontal gyrus	-10	52	36	4.14	3.94
	3466	Left STS/STG; middle occipital gyrus	-40	-52	18	4.12	3.93
	380	Right STS/STG	50	-16	-6	4.06	3.87
	978	Bilateral posterior-medial frontal lobe; bilateral mid- cingulate cortex	4	20	50	4.03	3.85
	459	Right IFG (pars triangularis);	44	30	6	3.92	3.75

RUNNING HEAD: Neural responses to talker gaze

		right insula						
	354	Bilateral lingual gyrus; bilateral calcarine gyrus	-14	-56	2	3.92	3.74	
	90	Left putamen	-32	-8	-6	3.61	3.47	
	289	Bilateral precuneus	4	-52	62	3.44	3.32	
	72	Right putamen	30	-4	0	3.44	3.32	
	175	Right precentral gyrus / central sulcus	56	-14	46	3.41	3.30	
	518	Left STS/STG	-54	-8	-2	3.40	3.28	
	84	Right superior parietal lobule	20	-56	54	2.93	2.85	
Direct >	2781	Bilateral calcarine gyrus; right superior occipital gyrus; left cerebellum (lobules IV-V)	24	-60	10	4.72	4.43	
Averted &								
Downward								
(T-test)	373	Left STG/STS; temporal pole; left Rolandic operculum; left postcentral gyrus	-62	-2	-6	4.56	4.30	
	69	Right cerebellum (lobules VIIIa, VI)	38	-46	-38	4.55	4.29	
	139	Right IFG (pars triangularis)	34	30	2	4.21	4.00	
	268	Right cerebellum (lobules IV- V, VI)	32	-38	-26	4.07	3.88	
	357	Left IFG (pars triangularis)	-42	22	0	3.70	3.55	
	125	Right insula	44	2	6	3.63	3.49	
	68	Left fusiform gyrus	-32	-42	-24	3.57	3.44	
	95	Right postcentral gyrus	60	-4	20	3.57	3.43	

RUNNING HEAD: Neural responses to talker gaze

	94	Right postcentral gyrus	42	-22	36	3.49	3.37
	70	Left superior occipital gyrus; right cuneus	-4	-84	46	3.46	3.34
	83	Right cerebellum (lobules VIIIa, VI)	8	-82	-28	3.30	3.19
	125	Left superior frontal gyrus	-28	60	14	3.22	3.12
Averted > Direct & Downward (T- test)	626	Right superior/inferior parietal lobule; right precuneus	20	-48	58	4.21	4.00
	77	Left inferior parietal lobule	-66	-32	36	3.81	3.65
	182	Right superior parietal lobule; right precuneus	16	-74	52	3.56	3.43
	102	Right superior frontal gyrus	36	-6	60	3.42	3.30
	117	Left superior frontal gyrus	-24	-4	54	3.26	3.16
Direct & Averted > Downward	4851	Bilateral calcarine gyrus; left precuneus	-10	-66	6	5.51	5.08
	314	Right insula	36	26	2	5.18	4.82
	327	Right superior frontal gyrus	26	-4	54	4.11	3.91
	425	Right post/precentral gyrus; right Rolandic operculum	60	-4	22	4.00	3.82
	536	Left mid-cingulate cortex; left posterior-medial frontal cortex	-6	2	44	3.99	3.81
	542	Right postcentral gyrus; right superior/inferior parietal lobule	40	-40	64	3.97	3.79
	759	Left post/precentral gyrus; left inferior parietal lobule	-34	-30	56	3.76	3.61

RUNNING HEAD: Neural responses to talker gaze

1	182	Right superior parietal lobule	-20	-52	60	3.60	3.47
2	130	Right central sulcus	28	-20	52	3.57	3.44
3	75	Left inferior parietal lobule	-36	-38	46	3.50	3.38
4	92	Left post/precentral gyrus	-56	-8	16	3.44	3.32
5	76	Left superior frontal gyrus	-26	-10	62	3.37	3.26
6	122	Left IFG (pars opercularis,	-44	10	28	3.36	3.25
7		triangularis); left precentral					
8		gyrus					
9	73	Left middle occipital gyrus	-42	-80	-2	3.29	3.18
10	93	Right inferior temporal gyrus;	52	-50	-16	3.28	3.17
11		right fusiform gyrus					
12	79	Right cerebellum (lobules VI;	24	-54	-24	3.25	3.15
13		IV-V); cerebellar vermis					
14	Conjunction	Left lingual gyrus	-16	-56	2	3.85	3.69
15	null: (Direct <	Left IFG (pars triangularis)	-42	24	-2	3.61	3.47
16	Averted &	Right insula	36	26	2	3.32	3.21
17	Downward) \cap	Right calcarine gyrus	16	-56	4	3.14	3.05
18	(3 channels > 2						
19	channels)						

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 2 Results of the laterality index calculations (using the LI toolbox in SPM; Wilke & Schmithorst, 2006) for main effects of Gaze Direction and Auditory Clarity, and for the directional contrasts of 3 Channels > 2 Channels and Direct > (Averted and Downward). All weighted means were significantly left lateralized (i.e. exceeding the conventional LI threshold of 0.2; Seghier et al., 2008).

Model	Contrast	Trimmed Mean	Min	Max	Weighted Mean
3 x 2 within-subjects ANOVA	Main Effect of Gaze Direction	0.4	0.18	0.65	0.62
	Main Effect of Auditory Clarity	0.44	0.29	0.62	0.5
	3 Channels > 2 Channels	0.25	0.03	0.52	0.42
	Direct > (Averted & Downward)	0.24	0.08	0.58	0.49
One-way within-subjects ANOVA (2 Channels)	Direct > (Averted & Downward)	0.25	0.2	0.38	0.44
One-way within-subjects	Direct > (Averted & Downward)	0.05	-0.01	0.09	-0.09

RUNNING HEAD: Neural responses to talker gaze

ANOVA (3 Channels)					
-----------------------	--	--	--	--	--

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 3 Results of the contrast Direct > (Averted and Downward) at each level of Auditory Clarity (2 Channels, 3 Channels), as tested within one-way within-subject ANOVAs with the factor Gaze Direction (Averted, Direct, Downward). All results are reported at a voxel height threshold of $p < .005$ (uncorrected), and a corrected cluster threshold of $p < .001$ (Slotnick et al., 2003). Coordinates are given in Montreal Neurological Institute (MNI) stereotactic space. WM = white matter

Auditory Clarity	No of Voxels	Region(s)	Peak Coordinate			F/T	Z
			x	y	z		
2 Channels	378	Left pre/postcentral gyrus	-28	-16	48	5.19	4.43
	221	Right calcarine gyrus, right lingual gyrus	26	-60	12	4.66	4.07
	530	Left calcarine gyrus	-12	-62	6	4.26	3.79

RUNNING HEAD: Neural responses to talker gaze

1	82	Right medial frontal WM	20	34	26	4.15	3.71
2	222	Right pre/postcentral gyrus	36	-28	66	4.11	3.68
3	79	Right IFG (pars triangularis),	38	32	2	4.11	3.68
4		insula					
5	127	Left temporal pole, superior	-54	14	-14	3.91	3.53
6		temporal gyrus, middle					
7		temporal gyrus					
8	192	L/R mid-cingulate cortex	-4	-2	38	3.54	3.24
9	167	Left IFG (pars triangularis)	-46	24	12	3.45	3.17
10	106	Left calcarine gyrus, middle	-4	-98	10	3.31	3.06
11		occipital gyrus					
12	3 Channels	Right cerebellum (Crus 1),	40	-48	-36	4.95	4.26
13		right fusiform gyrus					
14	301	Right Heschl's gyrus, Rolandic	54	-6	6	4.33	3.84
15		operculum					
16	116	Cerebellar vermis	0	-68	-26	3.79	3.44
17	160	Right calcarine gyrus	18	-54	4	3.72	3.38
18	96	Left cerebellum (Lobule IV-V,	-18	-46	-20	3.67	3.34
19		VI)					
20	93	Left Rolandic operculum,	-56	-6	8	3.56	3.26
21		postcentral gyrus					
22	77	Left superior temporal gyrus,	-62	-2	-8	3.48	3.19
23		temporal pole, middle temporal					
24		gyrus					
25	96	Right cerebellum (Lobule VI,	30	-54	-26	3.10	2.89

		IV-V)					
--	--	-------	--	--	--	--	--

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure legends

Figure 1. Example frames from the visual conditions used in the study. Please note that the Mouth Covered baseline was used in the behavioural pilot experiment only.

Figure 2. Plot of mean group accuracy (± 1 S.E.M.) on a test of sentence report, across the factors Visual Condition and Auditory Clarity.

Figure 3. Significant clusters showing a main effect of Auditory Clarity. Activations are shown at a voxel height threshold of $p < .005$ and a corrected cluster extent threshold of $p < .001$ (Slotnick et al., 2003). Plots show parameter estimates (± 1 S.E.M.) taken from 4mm-radius spherical regions-of-interest built around selected peak voxels (using the MarsBaR toolbox in SPM; Brett et al., 2002). Coordinates are given in Montreal Neurological Institute stereotactic space.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Figure 4. Significant clusters showing a main effect of Gaze Direction. Activations are shown at a voxel height threshold of $p < .005$ and a corrected cluster extent threshold of $p < .001$ (Slotnick et al., 2003). Plots show parameter estimates (± 1 S.E.M.) taken from 4mm-radius spherical regions-of-interest built around selected peak voxels (using the MarsBaR toolbox in SPM; Brett et al., 2002). Coordinates are given in Montreal Neurological Institute stereotactic space.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
Figure 5. Significant clusters showing a preferential response to Direct Gaze (compared with Averted Gaze and Downward Gaze; red). Also shown are regions showing a significant effect in the contrast 3 Channels $>$ 2 Channels (Auditory Clarity factor; green). Yellow shading indicates regions of overlap. Activations are shown at a voxel height threshold of $p < .005$ and a corrected cluster extent threshold of $p < .001$ (Slotnick et al., 2003). Coordinates are given in Montreal Neurological Institute stereotactic space.

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Figure 6. Significant clusters showing a preferential response to Direct and Averted Gaze (compared with Downward Gaze; blue). Also shown are regions showing a preferential response to Direct Gaze (compared with Averted Gaze and Downward Gaze; red). Magenta shading indicates regions of overlap. Activations are shown at a voxel height threshold of $p < .005$ and a corrected cluster extent threshold of $p < .001$ (Slotnick et al., 2003). Coordinates are given in Montreal Neurological Institute stereotactic space.

Figure 1
[Click here to download high resolution image](#)

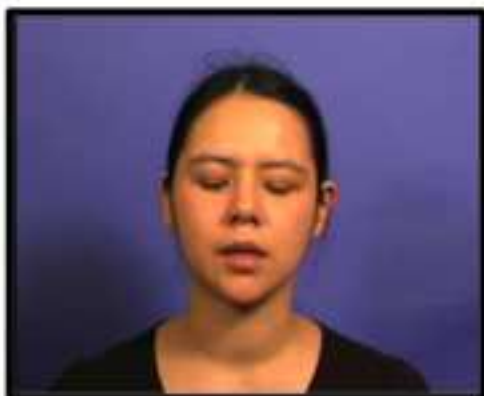
Experimental Conditions



Direct



Averted



Downward

Baselines



Eyes
Covered



Mouth
Covered

Figure 2
[Click here to download high resolution image](#)

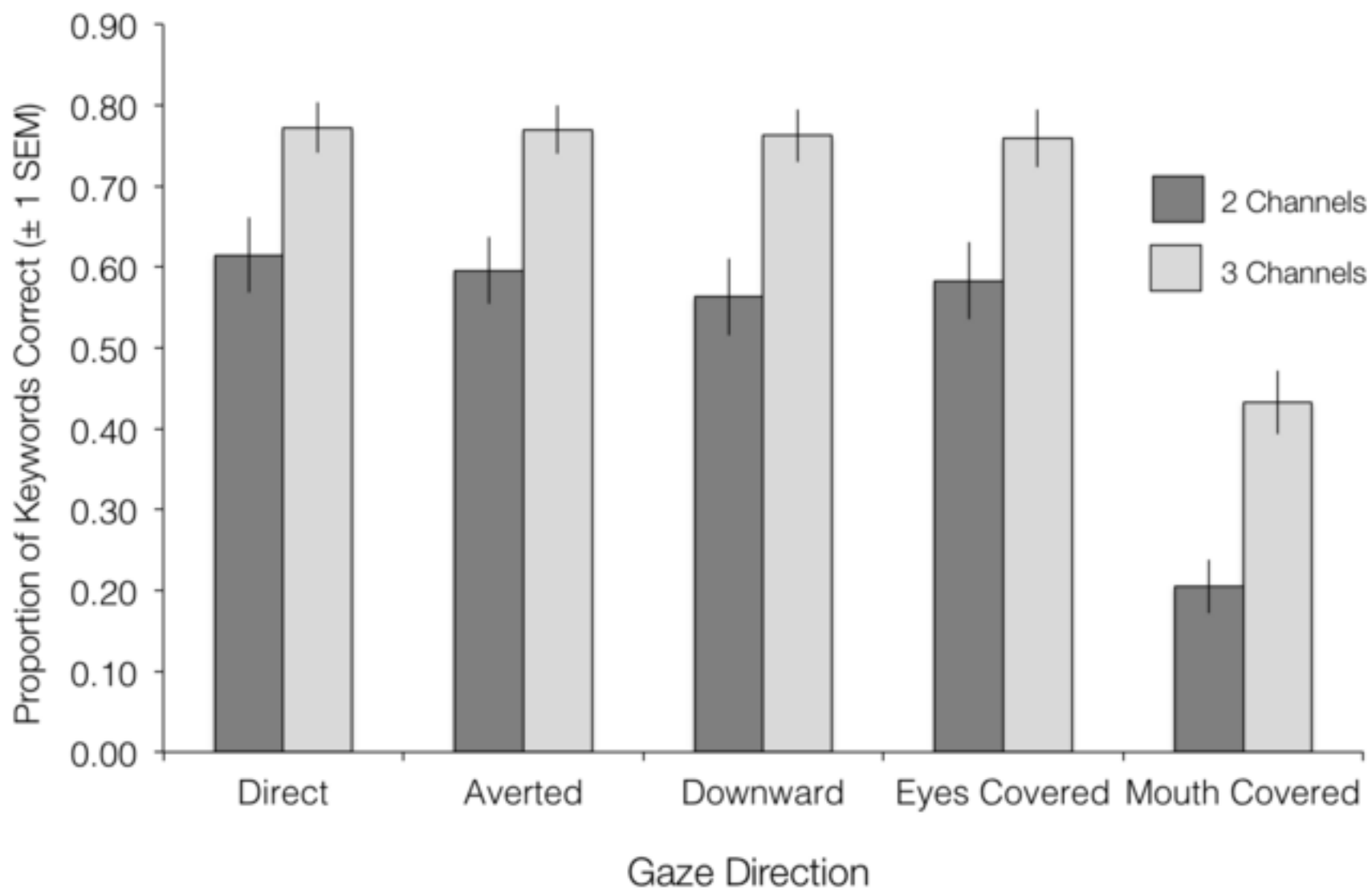


Figure 3
[Click here to download high resolution image](#)

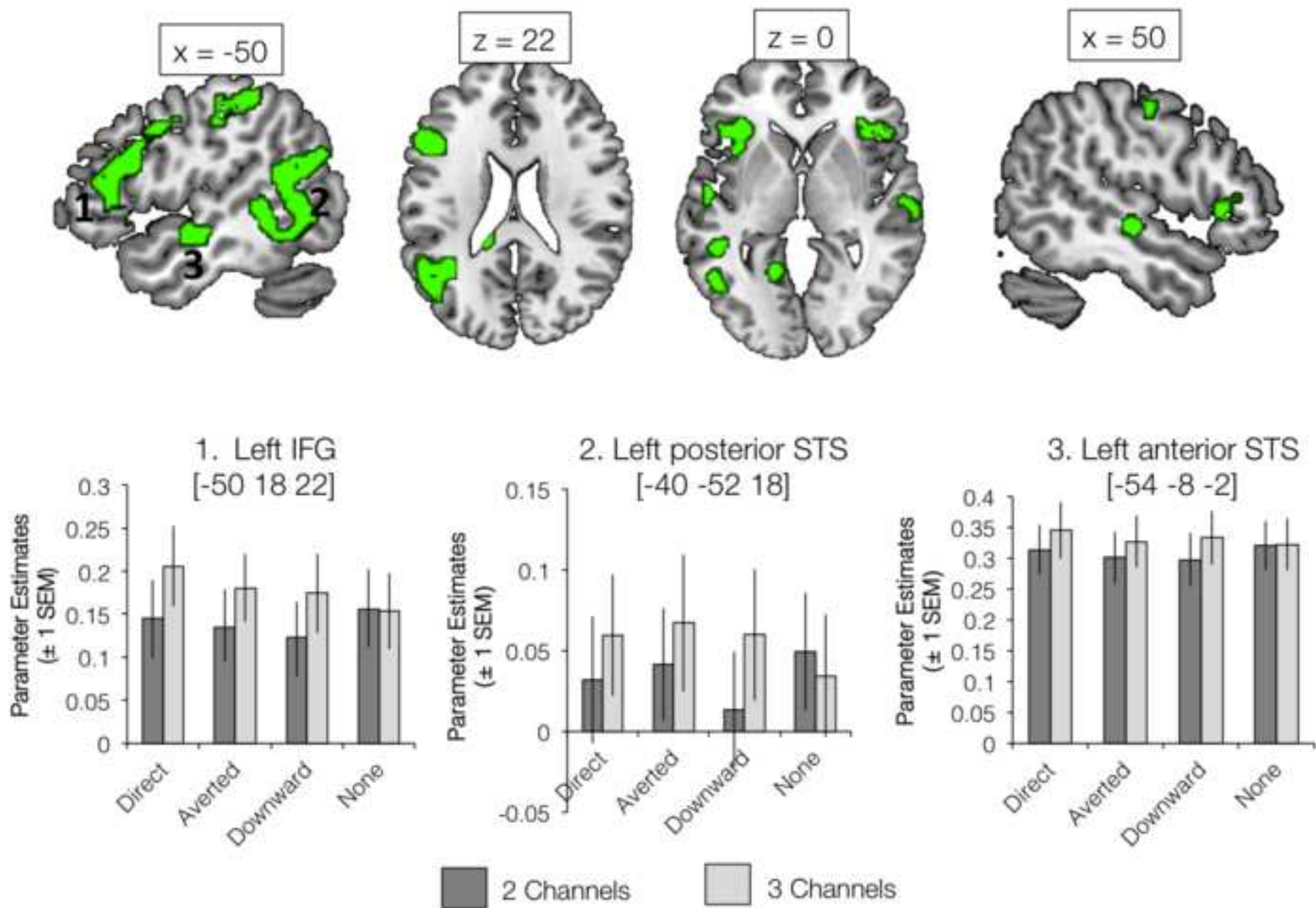


Figure 4
[Click here to download high resolution image](#)

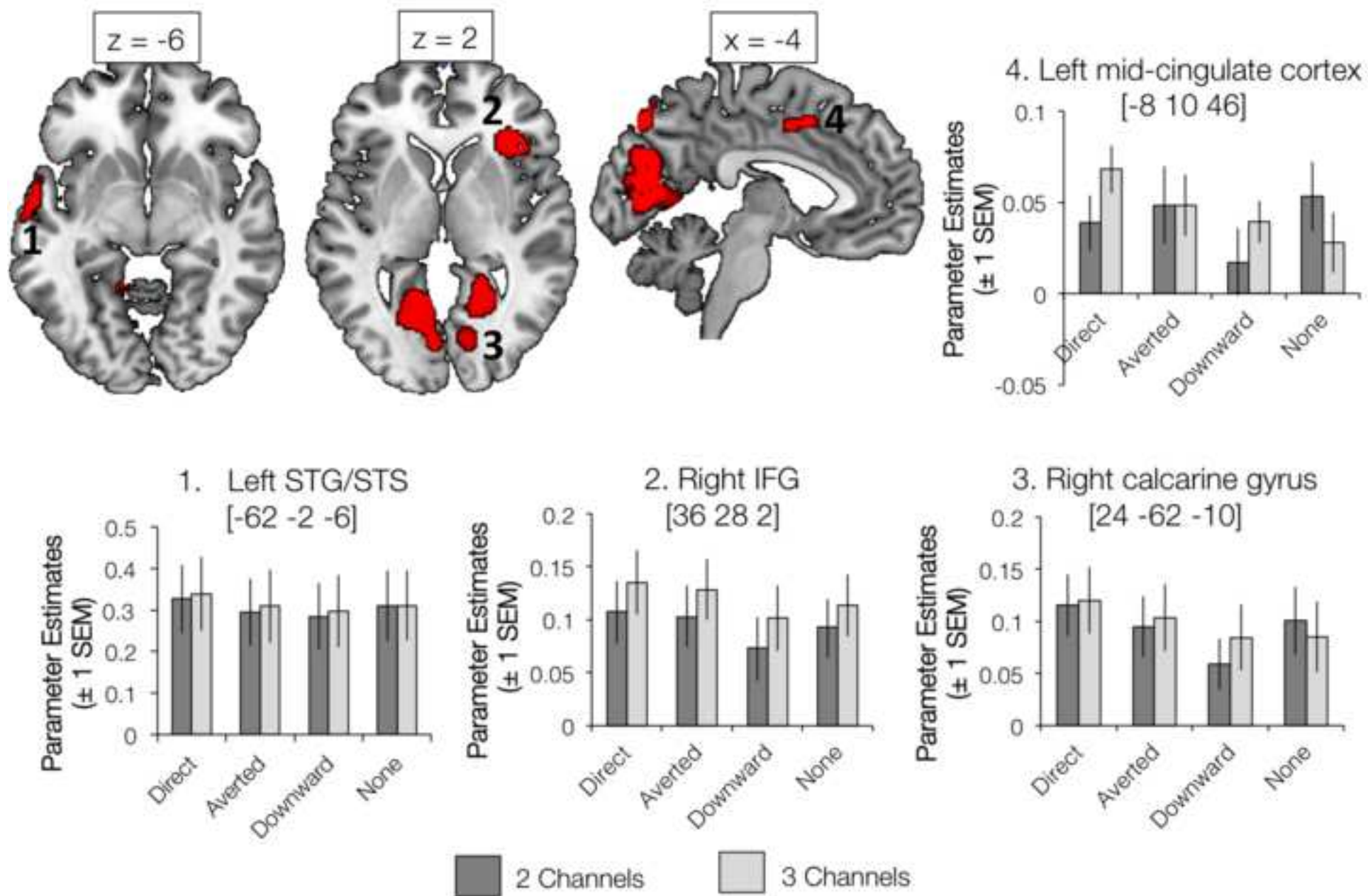


Figure 5
[Click here to download high resolution image](#)

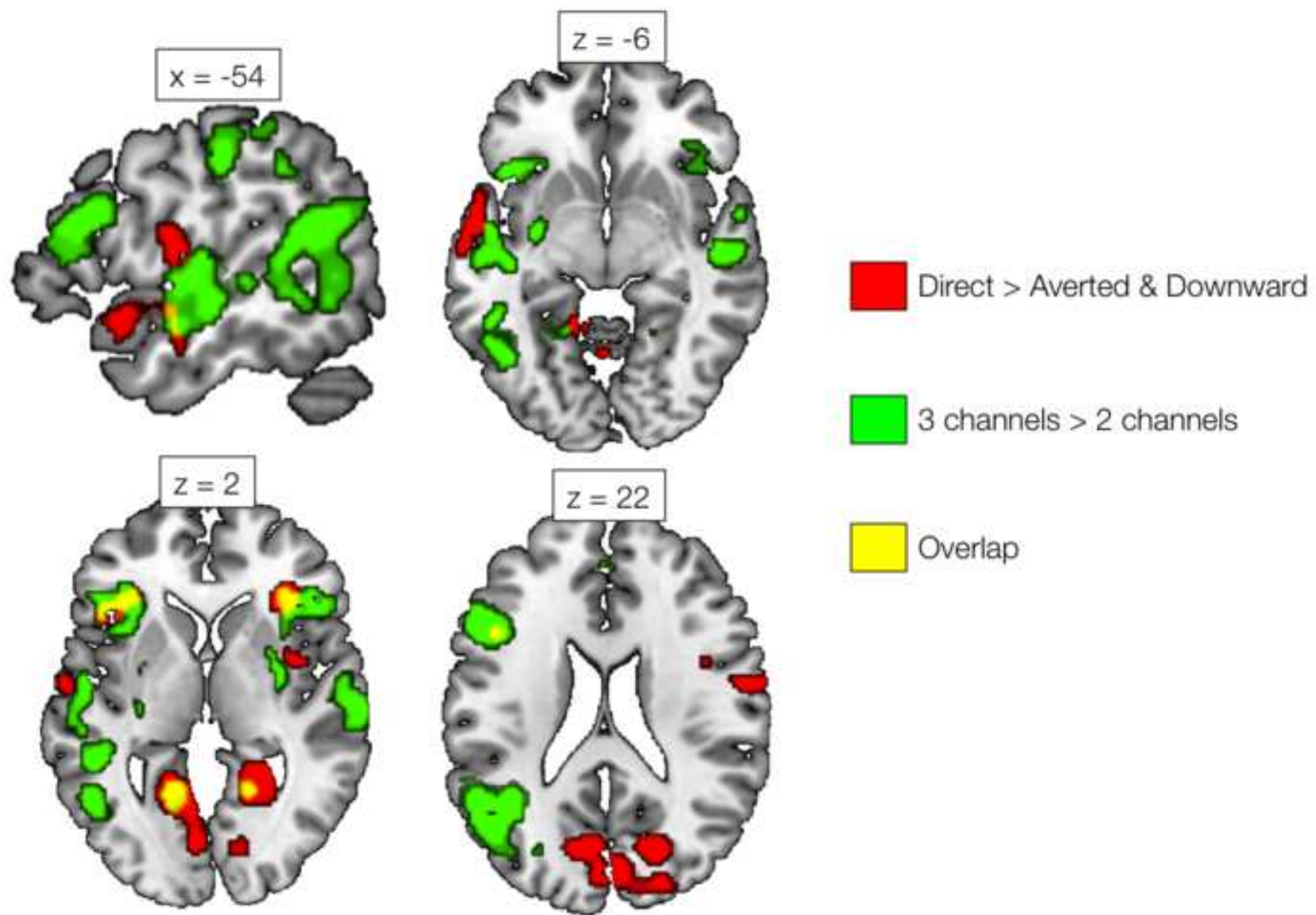
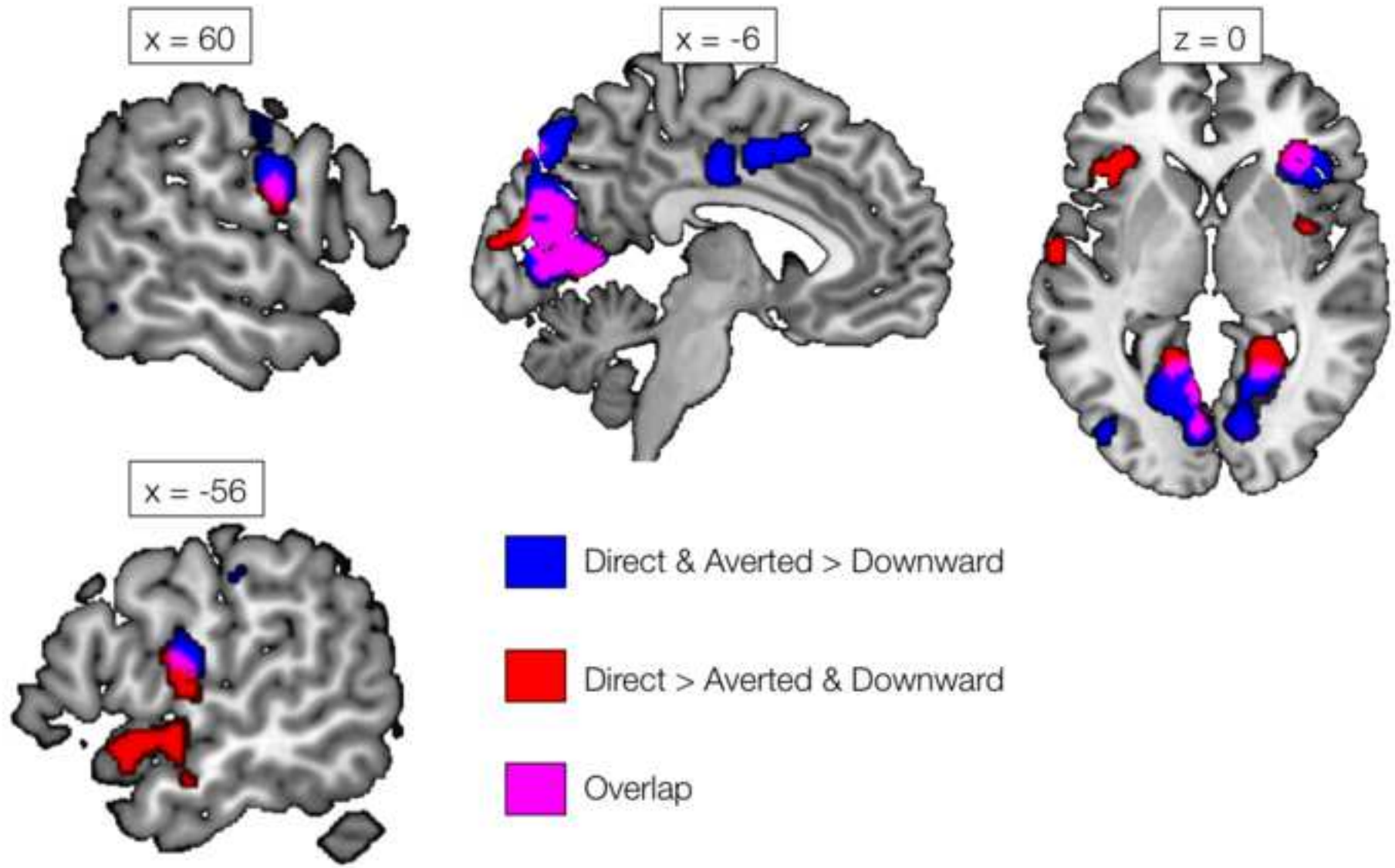


Figure 6
[Click here to download high resolution image](#)



Supplementary Material

[Click here to download Supplementary Material: McGettigan_Eyegaze_for_Neuropsychologia_R1_Supplemental_Materials.docx](#)