The role of the superior temporal gyrus in auditory feedback control of speech

Sophie Alexandra Louise Meekings

UCL

Thesis submitted for the degree of Doctor of Philosophy

I, Sophie Meekings, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

He gave man speech, and speech created thought,

Which is the measure of the universe.

— P.B. Shelley

What good are brains to a man? They only unsettle him.

—P.G. Wodehouse

ACKNOWLEDGEMENTS

Many thanks to my supervisors, Sophie Scott and Valerie Hazan, for their encouragement, support, and advice. I feel very lucky to have spent three years in Sophie's lab: not only is it a brilliant place to conduct research, but working there has been enormous fun, too.

Several of the experiments in this thesis required a lot of manpower to carry out, and I'm very grateful to the people who gave up their time (including evenings and weekends) to help me with testing: Nadine Lavan, Kyle Jasmin, Cesar Lima, Saloni Krishnan, Chris Worth, Erman Misirlisoy, Sophie Scott (again) and Sam Evans. Sam also provided me with a lot of help and advice in my first year and really got me off to a flying start.

I also wouldn't have any data without my participants. I am especially grateful to the participants with atypical voice control who lent me their brains: they were a delight to work with and many of them are living proof that you don't have to speak fluently to be a great communicator.

More personal thank-yous: Simone Brownless helped me settle on a thesis title, and Emily Fei was nearly responsible for calling it 'Harry Potter and the Role of the STG…'. Thanks to my family: Daniel lent his time and his ears (which work better than mine) to me on several occasions. Teresa, Bruno, Julian, Karl, Mary and Audrey provided me with love and support and didn't ask me how my thesis was going *too* often. Most of all, thanks to Llewe Gore for telling me when to stop working and relax (and occasionally making soothing whale noises to facilitate this. OooOOOOOOoooOOO).

Finally, my cats, Alfie & Marlowe, attempted to contribute to this thesis many times by walking across the keyboard, frequently covered in mud. Thanks, guys. I *think* I found all their insertions, but please blame any typos on my fuzzy editors.

THESIS ABSTRACT

Modern, biologically plausible models of speech production suggest that the superior temporal gyrus (STG) acts as a feedback monitor during speech production. This thesis investigates the role of the STG during speech production in three groups that have been hypothesized to use auditory feedback in differing ways: typical speakers, people who stammer, and a stroke patient. Because accurate speech production in most conversational settings can be accomplished without recourse to checking auditory feedback, it is necessary to introduce an external 'error', or feedback perturbation, to ensure that feedback control is being used. Here, masking noise was used as an ecologically valid perturbation that reliably prompts vocal adaptation.

An activation likelihood estimation meta-analysis showed that feedback perturbation is generally associated with bilateral STG activation. This was supported by a lesion study of a patient with left-sided stroke that suggested a link between temporal cortex infarct and an abnormal response to feedback perturbation. However, a functional magnetic resonance imaging (fMRI) study of typical speakers' behavioural and neural responses to different types of masking noise found that activation in the STG was driven not by the availability of auditory feedback, but by the informational content of the masker. Finally, an fMRI study of people who stutter— whose disfluency is hypothesized to arise from an overreliance on auditory feedback— found that STG activation was greatest in fluency-enhancing conditions, rather than during stuttering.

In sum, while there is some evidence that the STG acts as a feedback monitor, this is limited to a subset of situations that involve auditory feedback. It is likely that feedback monitoring is not as central to speech communication as the previous literature might indicate. It is suggested that the concept of auditory 'error' should be reformulated to acknowledge different types of speech goals—acoustic, semantic, or phonemic.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Speaking is a complex process that integrates conceptual, linguistic, respiratory and articulatory systems to create a communicative signal. It is such an impressive technical feat that only a few non-human animals even possess the requisite anatomy to replicate human vocal sounds, while some levels of the speech communication signal, such as compositional syntax, are arguably still unattested in non-human populations (Hurford, 2004). Given that speaking is such a complicated act, it is unsurprising that we do not always manage to execute it flawlessly. We may get a word stuck on the tip of our tongues (Brown, 1991) or come out with the wrong word altogether (Freud, 1915); or we may get phonemes mixed up and, like the Rev. W.A. Spooner, find ourselves proclaiming, "The Lord is a shoving leopard" rather than the more theologically sound, 'loving shepherd' (Stemberger, 1990).

How do we fix things when something goes wrong during articulation? A speech act results in auditory and somatosensory feedback, but neural processing delays and the fact that speech is a rapidly changing movement mean that this arrives at the ear too late to be of much use in correcting our utterances. This problem is compounded by the fact that we are almost never communicating in an ideal acoustic environment, so quite apart from our own errors there are other acoustic disturbances that may impede our attempts to communicate. Nevertheless, behavioural evidence (Cooke & Lu, 2010) suggests that humans are remarkably adept at making subtle, rapid vocal changes to compensate for challenging acoustic environments (for example, traffic noise, or competing

conversations). Although we cannot rely on sensory feedback all the time, the fact that we can adjust our voices in this way implies that we do use it as a source of information when producing our utterances. This thesis takes a critical look at the concept of auditory feedback control as currently formulated by neuroanatomical models of speech production. These models have implicated bilateral superior temporal gyri as a critical site for the processing of speech error and auditory feedback. The chapters that follow describe a series of investigations into three questions prompted by these models:

1. Is the STG reliably activated by feedback perturbation?

2. Do lesions involving the STG result in problems with feedback control?

3. Do people with a hypothesized impairment in feedback control display anomalous STG activation?

The answers to these specific questions will hopefully illuminate the following, broader and more philosophical questions: How central is feedback control to speech production? And what should we be defining as a 'speech error'?

The studies described in the chapters that follow deal with three distinct populations: typical speakers, people who stammer, and people with expressive aphasia. Issues relating to each specific population and experimental paradigm are therefore discussed in the introduction to the relevant chapter. The purpose of this chapter is to deliver an overview of the central concepts of the thesis. What follows is a discussion of the neural and psycholinguistic models of speech production, and how these have been integrated into the two computational models that serve as a starting point for the experiments described in this thesis. Next, the role of feedback control and the STG within these

models is examined, followed by a brief introduction to the techniques that are available for investigating feedback control. The chapter ends with a recapitulation of the aims of the thesis, and an outline of the way that this work as a whole aims to address its aims.

1.1. Wernicke-Lichtheim-Geschwind and beyond: brain areas involved in speech production

In the 1800s the work of Marc Dax and Paul Broca revealed an apparent link between damage to the left inferior frontal gyrus (specifically the pars triangularis and pars opercularis, or Brodmann areas 44 and 45) and a speech production deficit characterized by effortful, agrammatic utterances (Buckingham, 2006). Subsequently, in 1874, Karl Wernicke described two patients with lesions to left posterior superior temporal gyrus, whose speech was fluent but nonsensical, and who also displayed problems with language comprehension (Wernicke, 1874). Based on this work, in 1885 Lichtheim proposed an integrated model of speech now known as the Wernicke-Lichtheim-Geschwind model (Anderson et al., 1999; Lichtheim, 1885). In this model, Broca's area is responsible for expressive language (i.e. production of words), while Wernicke's area is responsible for receptive language (processing auditory input). Both are linked to each other via the arcuate fasciculus, and to a 'concept centre', suggested to lie in posterior medial temporal gyrus (Tranel, Damasio, & Damasio, 1997) which stores meanings.

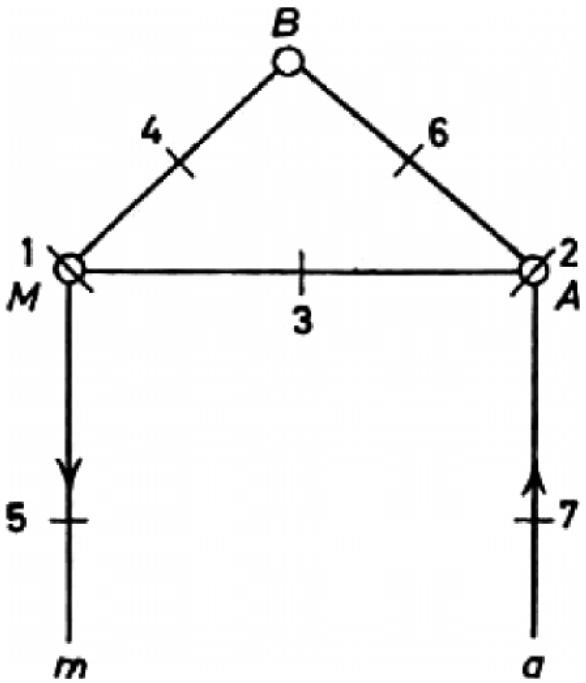


FIGURE 1: LICHTHEIM'S (1885) MODEL OF LANGUAGE ORGANISATION IN THE BRAIN.

In the diagram above, A represents the 'centre of auditory images' (Wernicke's area) which provides auditory input to the concept centre for interpretation, while the concept centre (B) delivers meanings to the 'centre of motor images', Broca's area (M), for articulation. Damage to any of the three regions or to the pathways connecting them, causes aphasia: Lesion 1 in the diagram above would cause Broca's aphasia; lesion 2 would cause Wernicke's aphasia, while damage to the white matter tract connecting them (lesion 3) results in 'conduction aphasia', in which comprehension and production are preserved but patients lose the ability to repeat words. Lesion 6, affecting the pathway from Wernicke's area to the concept centre, results in transcortical sensory aphasia, in which repetition skills are intact but comprehension is impaired. The corresponding impairment (lesion 4 in the diagram above) caused by damage to the pathway between Broca's area and the concept centre is called transcortical motor aphasia, and is marked

by a loss of spontaneous speech. Finally, damage to the concept centre itself should result in lost access to information about word meanings that patients can nevertheless hear and repeat (Lichtheim, 1885) .

Although the Wernicke-Lichtheim-Geschwind model has been extremely influential, in more recent years advances in linguistics on the one hand and in neuroscience on the other have demonstrated that the model inadequately describes the complexity of language production. For example, the division of language processing into 'receptive' and 'expressive' faculties is too broad to encapsulate accurately the linguistic distinctions between phonology, syntax and semantics that are now commonly recognised, let alone the many subdivisions of these categories (Levelt, Roelofs, & Meyer, 1999). Neuroanatomically, it has become clear that the different types of aphasia described by the model are not as straightforward or as easily dissociated as conceptualized. For example, patients with Broca's aphasia may make comprehension errors (Caramazza, Capitani, Rey, & Berndt, 2001), while some patients with Wernicke's aphasia have difficulty with speech production (Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980). Meanwhile, not all lesions to Wernicke's area result in Wernicke's aphasia, nor do lesions to Broca's area reliably result in Broca's aphasia (Bogen & Bogen, 1976; Mohr et al., 1978); this is discussed in more detail in chapter 4. Additionally, a recent reanalysis of the brains of Broca's original patients using high-resolution MRI (Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007) revealed that the damage to their brains not only included regions beyond the inferior frontal gyrus (such as the insula and superior longitudinal fasciculus) but also in both cases actually spared some of the region identified today as Broca's area (BA 44/45). Finally, conduction aphasia does not necessarily result from

damage to the arcuate fasciculus (Anderson et al., 1999), but arcuate fasciculus lesions can result in a more severe production deficit characterized by a complete loss of propositional speech (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004).

Our understanding of the anatomy and function of primary auditory cortex has also developed beyond Wernicke's area. In humans, primary auditory cortex is located in Heschl's gyrus, on the dorsal surface of the superior temporal gyrus. It is subdivided into three separate cytoarchitectonic regions, TE 1.0, TE 1.1 and TE 1.2 (Morosan, Schleicher, Amunts, & Zilles, 2005). Heschl's gyrus (HG) is bordered by the planum temporale on its posterior edge, and the planum polare anteriorally. There are separate pathways from posterior HG to posterior STG and from anterior HG to anterior STG (Upadhyay et al., 2008), and functional imaging has suggested that these represent different streams of processing: posterior auditory cortex is modulated by the position in which sounds are presented (Ahveninen et al., 2006; Van der Zwaag, Gentile, Gruetter, Spierer, & Clarke, 2011) while the anterior superior temporal sulcus (STS) responds to intelligible speech (Narain et al., 2003; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007; Scott, Blank, Rosen, & Wise, 2000). Modern accounts of speech processing have resolved this evidence into an integrated model featuring an antero-ventral 'what' stream of processing that is involved in identifying and extracting meaning from speech objects, while spatial processing is carried out by a postero-dorsal 'where' pathway (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) and sensorimotor integration occurs in the planum temporale (Griffiths & Warren, 2002; Warren, Wise, & Warren, 2005).

Beyond temporal cortex, subsequent research has linked a host of other neural regions to speech production (see Figure 2 for an example). Lesions in the superior tip of the

precentral gyrus of the insula result in apraxia of speech, a deficit in the ability to plan and execute the movements necessary for speech articulation (Dronkers, 1996), and the left anterior insula is active during articulation in neurotypical subjects (Indefrey & Levelt, 2000; Wise, Greene, Büchel, & Scott, 1999)

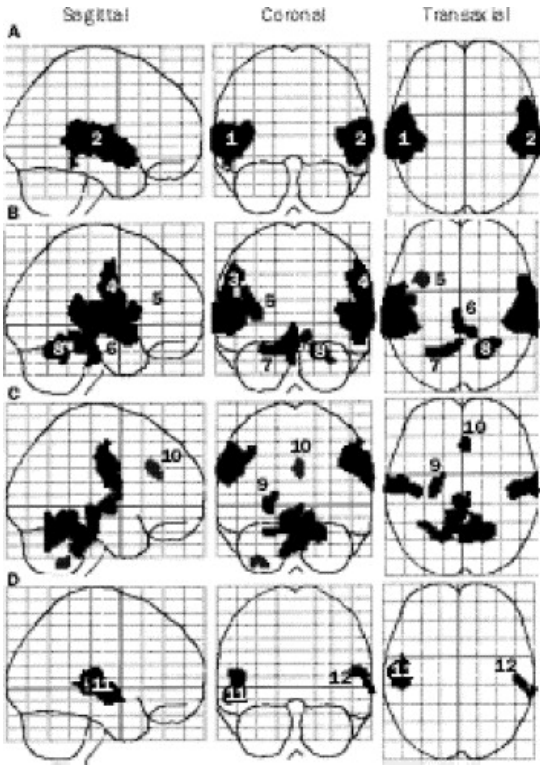


FIGURE 2: AN EXAMPLE OF BRAIN AREAS ASSOCIATED WITH SPEAKING AND LISTENING (FROM WISE ET AL 1999, REPRINTED WITH PERMISSION). A AND D SHOW ACTIVATION WHEN LISTENING IS CONTRASTED WITH WORD ANTICIPATION OR REPETITION; B AND C SHOW ACTIVATION WHEN WORD REPETITION IS CONTRASTED WITH LISTENING OR ANTICIPATION.

Articulation involves movement, and so the motor cortex is necessarily associated with speech production, particularly motor face cortex, which projects to the cranial nerves necessary for speech (Duffy, 2013). However, articulation is also associated with activation in the supplementary motor area (SMA), which is located in the superior frontal gyrus (Indefry & Levelt, 2000; Alario, Chainay, Lehericy, & Cohen, 2006). The SMA

has been linked to breath control during speech and vocalization (Murphy et al., 1997) and lesions to this area are associated with increased speech disfluency (Ziegler, Kilian, & Deger, 1997) as well as problems with speech initiation (Pai, 1999). The SMA and pre-SMA are involved in the selection and initiation of voluntary hand movements (Lau, Rogers, Haggard, & Passingham, 2004; Picard & Strick, 2001), and it is likely that this region has a comparable function during speech production.

Subcortically, the cerebellum has also been implicated in speech motor control (Nota & Honda, 2004; Riecker et al., 2005), and is involved in the fine motor control and temporal sequencing of overt utterances as well as inner speech (Ackermann, Mathiak, & Riecker, 2007); focal cerebellar lesions lead to a motor speech disorder called ataxic dysarthria, which is characterised by difficulty co-ordinating speech movements (Duffy, 2013). Lesions to the basal ganglia can also result in dysarthric speech (Pickett, Kuniholm, Protopapas, Friedman, & Lieberman, 1998). The basal ganglia are a group of five subcortical nuclei- the putamen, caudate nucleus, nucleus accumbens, globus pallidus and subthalamic nucleus, as well as the substantia nigra in the midbrain. They receive projections from cerebral cortex and return projections to cortex via the thalamus. The basal ganglia-thalamocortical loop forms part of a network (also including STG and premotor cortex) that is involved in the timing of self-paced motor sequences, and inhibits competing movements that might interfere with the desired action (Mink, 2003)

1.2. PSYCHOLINGUISTIC MODELS OF SPEECH PRODUCTION

While neurological models of speech production have addressed what happens during or after articulation, psycholinguistic models of speech production (Levelt, 1989; 1999)have

primarily focused on what happens prior to articulation. Levelt (1989) proposed two stages of word retrieval prior to articulation. First, the concept that the speaker wishes to express is encoded at the lemma level, which specifies modality-independent features of the word, such as grammatical class. The second stage involves retrieval of a lexeme representation- that is, the phonological aspects of the word. Conceptualizing word retrieval in this way offers a way to distinguish between words that differ in grammatical class or meaning (such as the insect 'fly' and the action, 'to fly'): the words have the same lexeme but different lemmas.

Levelt's model suggests that the two stages are discrete- that is, lemma selection must be complete before lexeme retrieval can begin. An alternative model, Dell's spreading activation account (Dell & Reich, 1981), allows for interaction between lemmas, lexemes and the semantic features of the word. That is, phonological processing can begin before lemma selection is complete, and can potentially feed back into lemma selection. Lemma selection arises from both top-down semantic activation and bottom-up phonological activation. This helps to account for 'mixed errors' in which words that are both semantically and phonologically similar to the intended word are selected- for example, 'rat' for 'cat' (Dell & Reich, 1981). Additionally, an interactive model helps account for situations in which talkers know the first phoneme of the word without knowing grammatical characteristics such as gender (Caramazza & Miozzo, 1997), which should not be possible if lemmas are fixed before lexeme selection begins.

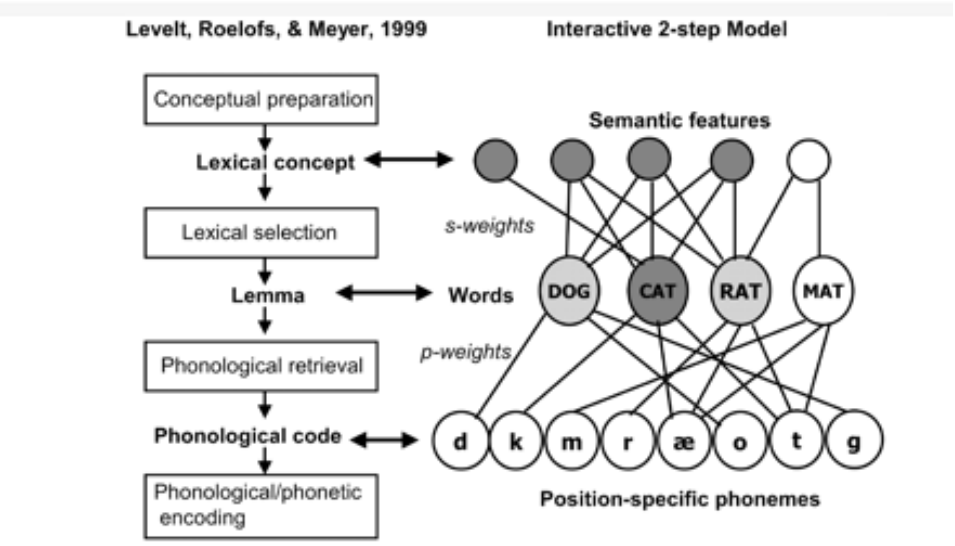


FIGURE 3: THE PROCESSING STEPS INVOLVED IN PRODUCING THE WORD 'CAT' IN LEVELT'S HIERARCHICAL MODEL COMPARED WITH DELL'S INTERACTIVE MODEL. (SCHWARTZ ET AL., 2009; REPRINTED WITH PERMISSION). S-WEIGHTS REPRESENT THE STRENGTH OF CONNECTIONS BETWEEN EACH SEMANTIC FEATURE AND THE VARIOUS WORDS (LEMMAS), WHILE P-WEIGHTS LINK WORDS TO THEIR CONSTITUENT PHONEMES.

Regardless of the specific model implemented, separating grammatical and phonological stages of word production helps account for quirks of speech production such as the tip-of-the-tongue phenomenon (Brown, 1991; Brown & McNeill, 1966) in which a talker knows, conceptually, the word they wish to produce but they are unable to produce it, despite knowing grammatical features of the word (Vigliocco, Antonini, & Garrett, 1997): the lemma has been activated but not the lexeme. Additionally, it can explain many of the error types found in speech production (Fromkin, 1971). For example, word substitutions tend to preserve grammatical class- that is, nouns swap for nouns and verbs swap for verbs (Garrett, 1992). Other errors result from phonological similarity between the target word and the error, including malapropisms, which occur at the word level (e.g. hysterical for historical) and spoonerisms, which occur at the phoneme level ('I must go

dye a beggar' for 'I must go buy a dagger'). These can be explained if the word is defined at the lemma stage but a mistake in word selection occurs at the lexeme level.

1.3. FEEDBACK AND FEEDFORWARD PROCESSES IN SPEECH CONTROL

The models described above outline some possible reasons behind the production of speech errors, but how are these errors detected and corrected? Levelt (1983) proposed a three-loop model of speech monitoring. 'Appropriateness monitoring' takes place during conceptualization, while an 'internal loop' checks the articulatory plan for errors after the lexeme has been selected, and finally the 'external loop' uses auditory feedback to fulfil postarticulatory monitoring. The idea that we listen to what we have said and check that it matches what we thought we were going to say is more technically known as the theory that speech is a *servosystem*- controlled by a mechanism that seeks to minimize error between intended and actual sensory feedback. This type of feedback control model, first proposed over 50 years ago (Fairbanks, 1954) is intuitive and persuasive enough that it has survived as a component in one of the most widely used modern computational speech production models, DIVA (Guenther, 2006). However, there are several logical and practical problems with a model based entirely on feedback control. Perhaps the most obvious is that by the time a sound has been produced, it is too late to use feedback from that utterance to correct itself. If we control our voices purely by listening to ourselves, we must first wait for the sound to arrive at our ears, then allow time for acoustic features such as pitch and spectral envelope to be estimated, and finally also factor in axon transmission times—meaning that it can take over 100ms after a sound is spoken for the information it contains to become available to the central nervous

system (Hickok, 2012). When you consider that on average humans produce speech at a rate of approximately 10 phonemes per second (Osser & Peng, 1964) it becomes apparent that not only does this processing delay render feedback largely useless in terms of real-time speech correction, but according to feedback control theory (Franklin, Powell, & Emami-Naeini, 2002) would cause a model based entirely on feedback to become unstable.

One approach to solving this problem is to propose that speech production consists of two subsystems, a feedback control system that monitors for errors, and a feedforward control that takes on the bulk of the work by informing speakers how to produce sounds. This is the approach taken by Guenther's (2006) DIVA model (Fig 4). The model consists of a network of "maps" (or sets of representations), each with a proposed anatomical location based on past neurophysiological research. A speech sound map defines auditory and somatosensory targets (in terms of syllables) and sends commands to motor cortex. Once the command has been executed, the feedback control system looks for discrepancies between feedback and sensory targets, generating an error signal that leads to an adjustment in the direction of the target. Because of the problems with using feedback to monitor speech in real time previously noted, in the DIVA model this system is primarily used only in development, to refine motor plans based on feedback about effective associations between motor commands and sensory outcomes; once these associations are learned, speakers can rely on the feedforward system to guide them through speech production. However, feedback remains the only way to catch errors if they occur.

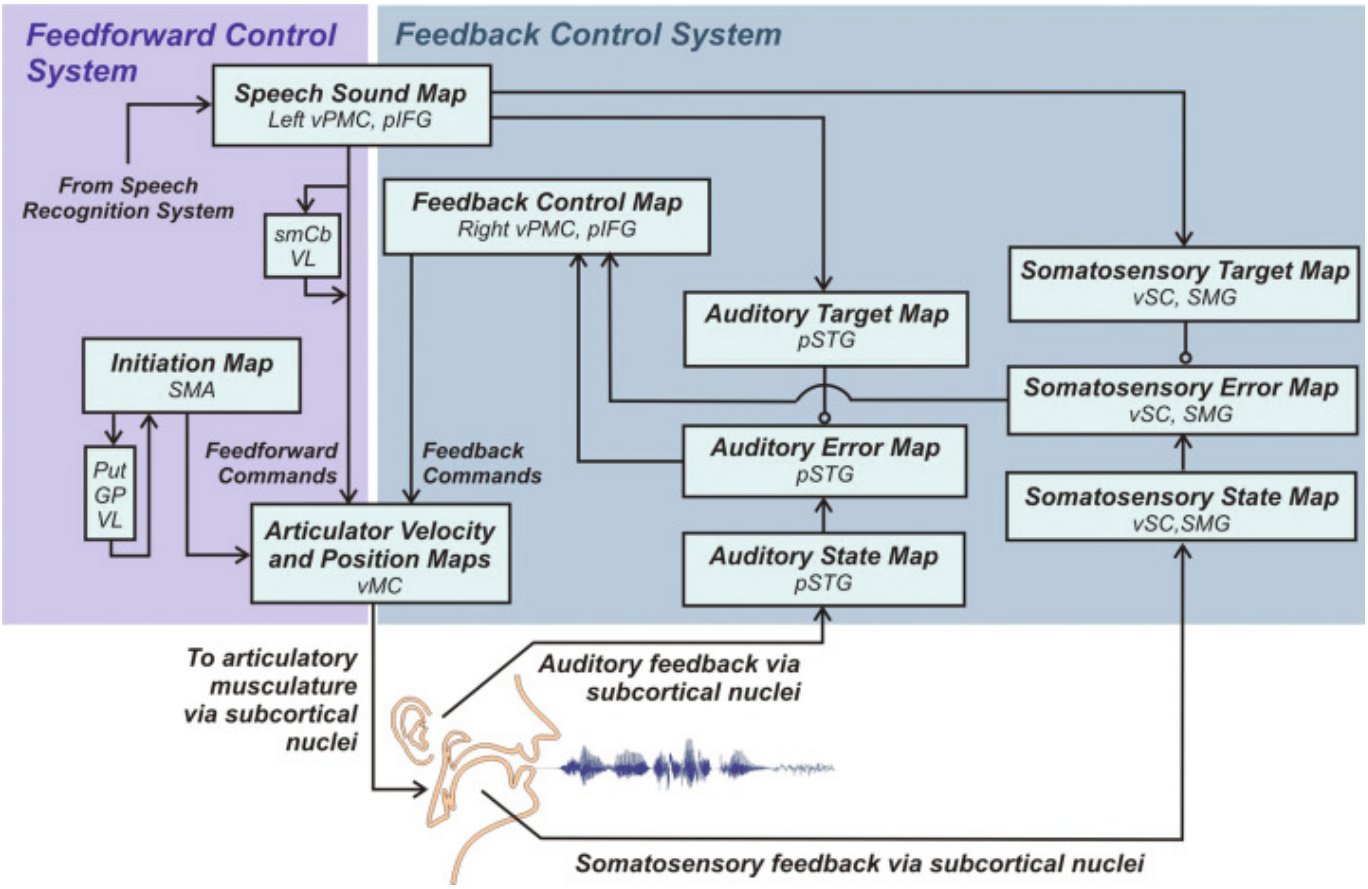


FIGURE 4: THE DIVA MODEL (GUENTHER & HICKOK, 2015)**. GP, GLOBUS PALLIDUS; PUT, PUTAMEN; PIFG, POSTERIOR INFERIOR FRONTAL GYRUS; PSTG, POSTERIOR SUPERIOR TEMPORAL GYRUS; SMCB, SUPERIOR MEDIAL CEREBELLUM; SMA, SUPPLEMENTARY MOTOR AREA; SMG, SUPRAMARGINAL GYRUS; VL, VENTRAL LATERAL NUCLEUS OF THE THALAMUS; VMC, VENTRAL MOTOR CORTEX; VPMC, VENTRAL PREMOTOR CORTEX; VSC, VENTRAL SOMATOSENSORY CORTEX.REPRODUCED WITH PERMISSION.**

This is somewhat problematic for cases where feedback is radically different from the syllable target—for example, when speech is masked by noise. In this case there is such a large mismatch between what the speaker hears and auditory goals that the compensatory adjustments made could in theory render speech unintelligible.

As an alternative, Hickok's (2012) Hierarchical State Feedback Control (HSFC) model (Fig. 5) proposes that feedback is compared not to auditory goals directly, but to an internal model of the predicted consequences of motor commands. This internal model contains an estimate of signal noise that helps to ameliorate the mismatch between feedback and prediction in busy acoustic environments. Here, similarly to DIVA, feedback is primarily used to update internal representations rather than to correct speech after articulation (although it can still be used for this purpose). However, in the HSFC model, outgoing motor commands are checked against the internal model, meaning that feedback forms a part of the selection process rather than simply serving to correct errors. Another difference between the HSFC model and DIVA is that Hickok's model incorporates a hierarchy where syllable goals are activated first and then project to a lower, articulatory feature cluster level. Auditory targets are defined at the syllable level, while motor targets fill in the fine, phoneme-level detail.

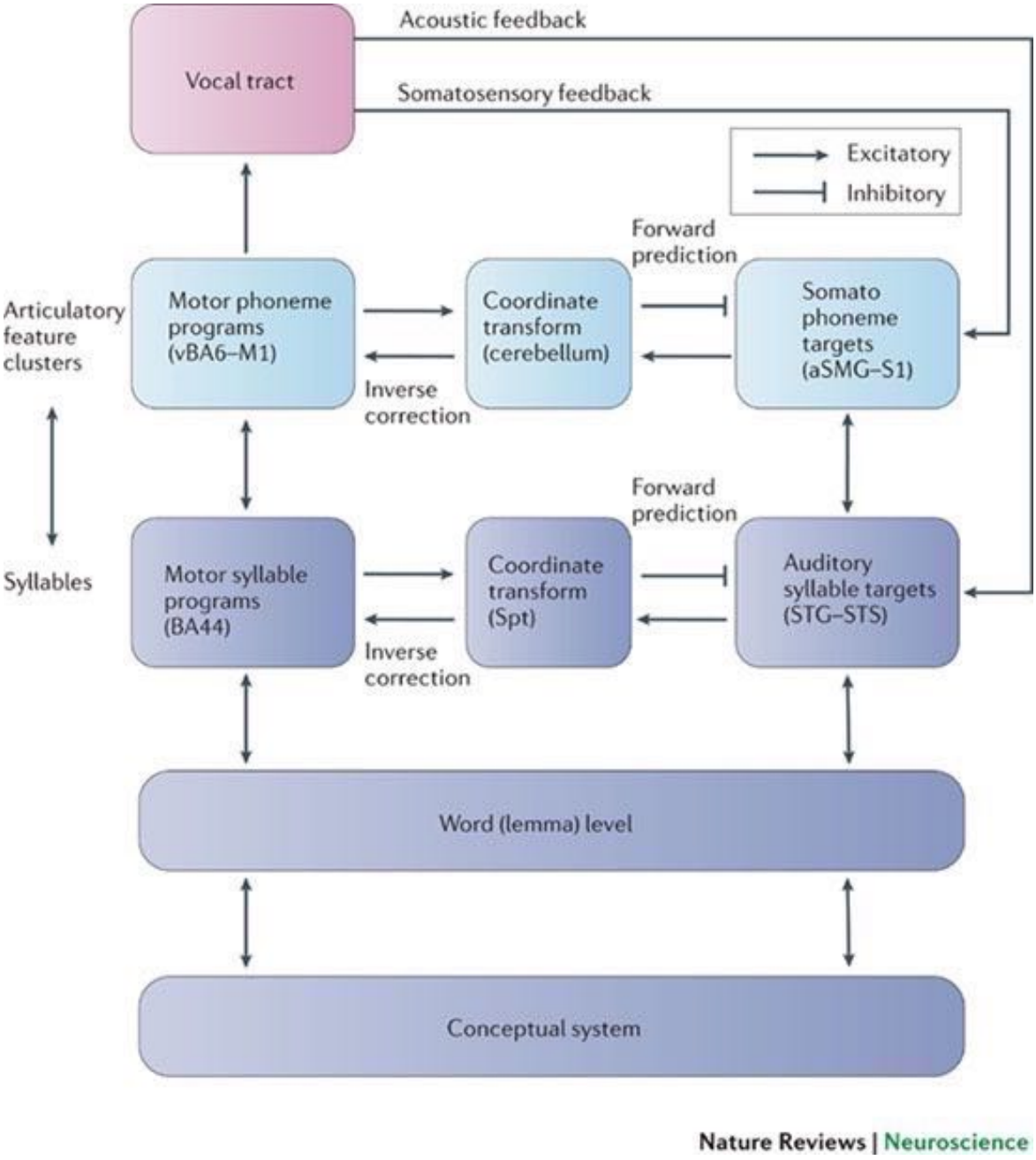


FIGURE 5: HIERARCHICAL STATE FEEDBACK MODEL (HICKOK, 2012). V: VENTRAL; A: ANTERIOR; BA: BRODMANN AREA. REPRODUCED WITH PERMISSION.

The model begins when a conceptual representation is activated and triggers the activation of a word representation (lemma). This projects to both auditory/phonological and motor components simultaneously at the syllable level, before cascading to the

phoneme level. The motor and sensory systems project to each other, with an inhibitory motor-to-sensory target signal and an excitatory sensory-to-motor signal, which is intended to activate the associated motor programme. The excitatory signal is intended as a correction signal that, if unneeded, is essentially cancelled out by the inhibitory signal. If, however, the wrong motor programme is activated, the motor-to-sensory signal will be misdirected and the excitatory impulse will not be inhibited, leading to correction of the motor syllable programmes in BA44.

Although the models are conceptually quite different, it can be seen by this that they predict broadly the same end result in terms of neural activation. Crucially, for our purposes, both models predict that superior temporal cortex (STG/STS) encodes auditory targets, and therefore activation in this area during speaking is driven by the amount of mismatch between that target and what was actually produced (Guenther, 2006) or predicted (Hickok, 2012). That is, in cases where feedback is normal we ought to see suppression in superior temporal cortex during normal speech compared to listening to another's voice. However, when targets are not met (because feedback is in some way inaccurate), this mismatch triggers activation in that area; the greater the mismatch, the stronger the activation.

Somewhat unsurprisingly (since the neural correlates of each model's components have been posited based on past research), there is considerable evidence from neurophysiological studies in human and non-human primates that activity in superior temporal cortex associated with passive listening is attenuated when subjects speak alone. Eliades and Wang (2003) used single-cell cortical recordings to investigate the response to vocalizations in primate auditory cortex, finding that most cells

demonstrated suppressed activity during self-initiated vocalizations, with suppression beginning approximately 220ms before vocal onset. A smaller population of neurons increased their activity during the production of speech, but this activation did not occur until after the onset of vocalization. In humans, the vocalization-induced suppression response has been demonstrated in the superior temporal gyrus (STG) using cortical recording (Creutzfeldt, Ojemann, & Lettich, 1989; Flinker et al., 2010), fMRI (Agnew, McGettigan, Banks, & Scott, 2013), PET (Wise et al., 1999) and MEG (Houde, Nagarajan, Sekihara, & Merzenich, 2002).

Neurally and physically, there is a clear distinction between the way that self-generated vocalizations are processed compared to externally produced sounds, with self-produced speech resulting in disengagement of systems involved in passive listening, usually occurring prior to the onset of vocalization. Physiologically, speech production triggers contraction of the tensor tympani and stapedius muscles in the middle ear, a reflex that physically attenuates sound by reducing the efficiency of the middle ear as a sound transmitter. This means that when people speak, they perceive their voice as quieter than an external sound with equivalent intensity. Behaviourally, this is reflected in the fact that that speakers consider themselves to have doubled their vocal level with only 2/3 the increase in sound pressure that it would take for them to consider an external sound to be twice as loud (Lane, Tranel, & Sisson, 1970). Although hearing other loud sounds can also prompt middle ear muscle contractions, in this case contractions are initiated between 90 and 300ms after sound onset. By contrast, when the response is elicited by self-produced vocalizations it always precedes (by up to 300ms) or happens simultaneously with sound onset (Salomon & Starr, 1963). This indicates that

vocalization-induced middle ear contractions occur not as a consequence of the physical properties of the sound produced, but rather in preparation for the production of speech. This, together with the neural evidence described above that suppression in primary auditory cortex begins prior to the onset of articulation, is hard to interpret with relation to the claim that suppression results from neural cancellation resulting from accurate feedback, since the suppression is initiated before any commands have been executed. It has been suggested that this response is not related to error monitoring, but rather a means of distinguishing self-produced vocalizations from external sounds (Scott, 2012), analogous to the suppression of cortical responses to self-generated touch compared to the touch of others (Blakemore, Wolpert, & Frith, 1998).

Schizophrenic patients are less able to distinguish both self-generated touch from other-produced touch (Blakemore, Smith, Steel, Johnstone, & Frith, 2000) and self-generated speech from other-produced speech (Johns et al, 2001), indicating that the two systems may operate on similar principles. It has been suggested (Frith & Done, 1988) that auditory hallucinations experienced by such patients are caused by an inability to tell when sounds are self-initiated. Functionally, this deficit has been linked to abnormal activity in regions associated with verbal self-monitoring, including cerebellum and temporal cortex (Shergill, Bullmore, Simmons, Murray, & McGuire, 2014), implying that the perceptual system plays a key role in helping speakers attribute agency to their own utterances.

1.4. ALTERED AUDITORY FEEDBACK

Since the evidence for the role of the STG in processing feedback during accurate speech is somewhat ambiguous, it may be more informative to look at what happens when feedback is off-target. If STG encodes auditory error, then activation in STG ought to correlate with the degree of mismatch between feedback and prediction (or auditory target). It is, therefore, important to include a listening baseline in any study of altered feedback, to exclude the possibility that any activation seen is simply the result of hearing an unusual sound. Whilst models predict that superior temporal cortex activates more for speech in altered feedback than for normal feedback, recall that it has already been demonstrated (and indeed is an integral feature of both models) that exactly the same effect is observed when listening is compared to normal feedback. A true prediction mismatch response ought to be greater in altered feedback compared to both unaltered feedback and listening to the sound without articulation. There are three common types of feedback manipulation used by these studies: delayed auditory feedback (DAF), frequency altered feedback (FAF) and masking sound (MASK). A brief introduction to these techniques is given below, while Chapter 3 details a systematic review and meta-analysis of functional imaging studies that have used these types of feedback perturbation.

Speaking in delayed auditory feedback, a situation analogous to talking on a phone with a bad connection, has been shown to induce dysfluent speech in typical speakers, which has been interpreted as demonstrating the importance of accurate auditory feedback to speech production (Yates, 1963). However, research into the nuances of the behavioural response to delayed auditory feedback suggests that the disruption caused by DAF may

result from more general problems with action timing, rather than with speech feedback. The delay that causes maximal disruption to speech is typically 200ms (Black, 1951)--the duration of a syllable. This has led some authors (Howell & Powell, 1987) to suggest that the disruption caused by delayed auditory feedback is due to lower-level timing processes associated with trying to execute any rhythmic gesture, such as clapping.

A different research technique involves manipulating the frequency of speakers' voices in real time (frequency altered feedback, or FAF). Here, the mismatch between feedback and prediction is experimentally induced by asking the subject to speak into a microphone, shifting the frequency of the recorded speech, and then playing this back to the participant via headphones. Behavioural evidence suggests that subjects can compensate for the change by shifting their voice in the opposite direction to the altered feedback, even when the frequency shift happens just at the phoneme level- for example, when adjustments are made to individual formant frequencies rather than to the frequency of the utterance as a whole (Tourville, Reilly, & Guenther, 2008). However, not all participants display a compensatory response to auditory feedback, and some show a preference for a particular feedback modality, responding to somatosensory feedback perturbation, but not auditory feedback perturbation, and vice versa (Lametti, Nasir, & Ostry, 2012).

An additional drawback of the techniques detailed above is that, even allowing for the intentional manipulation, the subject is not perceiving their voice in exactly the same way that they would normally experience feedback. Self-produced speech is very different spatially (because it comes from your mouth) and acoustically (because it is conveyed through bone conduction) to a recording of your voice played back at the ear, through air

conduction. Given that the experience of hearing your voice played back to you is so different from the experience of hearing it normally, it is even possible that subjects might not perceive altered feedback as self-produced, instead treating it as an external sound. In this thesis, perceptual experience in all three experiments was manipulated using masking sounds. Masking sound does not alter any of the acoustic or somatosensory properties of self-produced speech. Instead, it alters perception of auditory feedback by providing a competing sound that may obscure or distract from the feedback signal.

In chapter 5, we distinguish between two properties of masking sound: informational and energetic masking potential. The energetic potential of a masker determines how effectively the masker's acoustic properties interact with those of the signal, resulting in overlapping patterns of excitation at the periphery of the auditory system over time (Festen & Plomp, 1990; Stone, Füllgrabe, & Moore, 2012). Thus, the energetic masking potential of a noise is determined by acoustic properties such as its frequency spectrum and intensity relative to the signal (Brungart, 2001), and properties of random amplitude fluctuations (Stone, Füllgrabe, Mackinnon, & Moore, 2011). Meanwhile, masking properties that cannot be explained by the energetic properties of the masking noise are described as its informational masking potential (Shinn-Cunningham, 2008). An informational masker creates competition for more central cognitive resources, often because the sound contains some kind of salient or meaningful content that could distract the listener (Carhart, 1969). Mattys et al. (Mattys, Brooks, & Cooke, 2009) suggested three component processes involved in informational masking: competing attention, reflecting the effort required to segregate the target from the masker; intelligibility, which creates lexical-semantic competition between target and maker; and increased cognitive load

associated with processing the masker. From this it can be seen that, whilst speech is the most commonly discussed informational masker, sounds do not need to have semantic content to provide 'information' in the sense used here. Thus, when carrying on a conversation at the famous annual speech science cocktail party (Pollack & Pickett, 1957) informational masking is provided not just by the juicy gossip happening behind you, but potentially also by the wail of the police siren telling you it might be time to drop your drink and scarper.

Functional imaging studies of speech perception have established that informational and energetic maskers activate different neural systems. Consistent with the notion that informational masking is associated with greater competition for central resources, trying to understand speech masked by another talker results in bilateral activation of the superior temporal gyrus (Scott, Rosen, Beaman, Davis, & Wise, 2009). In contrast, listening to speech against an energetic masker is associated with activations in prefrontal and posterior parietal cortex, which implies an increase in attentional rather than linguistic resources (Scott, Rosen, Wickham, & Wise, 2004).

1.5. OUTLINE OF THE THESIS

The idea of the STG as an error monitor has dominated speech production research for the past decade, and has been used to explain both typical speech production and speech disorders. Here, we investigate the claims made about the superior temporal gyrus and its role in feedback control of speech production.

Chapters 3 and 5 investigate the question: **Is the STG reliably activated by feedback perturbation?** In chapter 3, a systematic review of the neuroimaging literature is conducted, followed by an ALE meta-analysis of the co-ordinates reported for feedback perturbation contrasted with non-perturbed speech. This found that there was significant convergence among reported co-ordinates for perturbed compared to unperturbed speech, but only a handful of studies included an appropriate listening control condition, limiting the inferences that can be drawn from this. Chapter 5 delves deeper into the question by looking at whether STG activation is modulated by how well you can hear yourself, or by other properties of the environment. The results demonstrate that the STG is modulated significantly more by informational properties of the masking sound than by how effectively the sound perturbs auditory feedback by obscuring the talker's voice.

Chapter 4 is a stroke study addressing the question: **Do lesions involving the STG result in problems with feedback control?** The case of a man who suffered a left-sided stroke affecting the temporal lobe is discussed. This patient reported experiencing attenuated auditory feedback and was not found to have any other hearing or sound perception

problems. When he spoke in noise, he over-compensated for maskers compared to a control group, possibly supporting a role for temporal cortex in feedback perception.

Chapter 6 asks, **Do people with a hypothesised impairment in feedback control display anomalous STG activation?** People who stammer (PWS) are thought to rely on feedback control more than typical speakers do. We collected behavioural and neural data on how PWS are affected by two types of altered feedback, synchronous speech and masking sound, compared to speaking in quiet. Speaking in noise and synchronous speech are shown to recruit clearly dissociable networks in the brain, despite producing comparable behavioural responses. Additionally, participants did not display the speaking-induced suppression response in STG characteristic of neurotypical speech production. A summary of the aims and results of the thesis is given in Chapter 7, together with comments on future direction

CHAPTER 2: METHODS

2.1. INTRODUCTION TO MAGNETIC RESONANCE IMAGING

The human body is mostly made of water. Since different types of tissue have different concentrations of water, we can differentiate between tissue types just by looking at the water content. This is what an MRI scanner does: it creates images of the body's soft tissues based on the amount of water they contain. It can do this because the hydrogen protons within each water molecule possess the nuclear magnetic resonance (NMR) property. That is, it has both a magnetic moment (or 'spin') and an angular momentum. The spin creates a small electrical current and a magnetic source on the surface of the nucleus. In the absence of a magnetic field, the spin axes of the protons are oriented randomly, and on average cancel each other out. When in the presence of a strong magnetic field like that created by the MRI scanner, most of the proton axes align in the direction of the magnetic field (parallel to it), while a minority remain oriented in the opposite (antiparallel) direction. Overall the net magnetization is longitudinal. This is a low-energy state. If an excitation pulse at the resonant frequency of the protons is applied, some of the protons will absorb energy from it and change to an antiparallel orientation- a high-energy state. The angle to which the net magnetization is tipped by the application of the RF pulse is called the flip angle. Once the pulse is switched off, the spins release energy, which is received by the coil, as they return to the low energy state. The time it takes for the net magnetization to return to the longitudinal orientation is called the T1 relaxation time. T1-weighted images are commonly used for structural

scans upon which the functional activation is superimposed. Tissues with different amounts of water (and therefore different numbers of spins) have corresponding T1 relaxation times- cerebrospinal fluid (CSF) is very slow, white matter is very fast. The faster a tissue's T1 relaxation time, the brighter it appears on a T1-weighted image: white matter is very bright, grey matter slightly darker, and CSF is very dark. The reverse is true for T2-weighted images, which measure the rate of the decay of transverse magnetization- that is, the rate at which the spins fall out of phase with each other through spin-spin interactions (in which they affect each other by their individual oscillating magnetic fields). T2-weighted images are most commonly used to aid medical diagnoses, as most pathologies are associated with an increase in water content, which shows brightly on T2-weighted images. Functional images are obtained by using T2* weighting, which includes the dephasing caused by magnetic field inhomogeneities and susceptibility effects as well as particle interactions.

2.2. FUNCTIONAL MRI

Functional MRI (fMRI) uses T2* weighting to measure changes in the magnetic field associated with distortions caused by blood flow. The brain cannot store oxygen or glucose, so energy is replenished via the blood supply. If an area is more active than usual, more energy is needed to resupply the cells, and so blood flow through that area will also increase.

Blood carries oxygen in haemoglobin molecules. Each molecule of haemoglobin has 4 iron molecules to which oxygen can attach. Once the oxygen is detached, the iron molecules are exposed: oxyhaemoglobin (di-magnetic) is converted to deoxyhaemoglobin (para-

magnetic), distorting the magnetic field. This means that nearby protons experience different field strengths and will precess at different frequencies. This causes them to lose phase with each other more quickly, resulting in a shorter T2*. The distortion is measured to determine the concentration of deoxyhaemoglobin present in blood. This is the BOLD signal (blood oxygen-level dependent signal). The assumption underlying fMRI research is that if performing a task is associated with an increase in the BOLD signal in some part of the brain, that region is involved in the cognitive functions underlying the task. Although there is considerable variability in the shape of the BOLD signal from person to person, and also between areas in the brain, it can generally be identified by a characteristic pattern (Figure 6 below), which consists of a brief dip in oxygen levels followed by a steady increase that peaks at around six seconds after the stimulus onset. The signal then undershoots, dropping below pre-stimulus levels, before returning to baseline around 20-30s after stimulus onset.

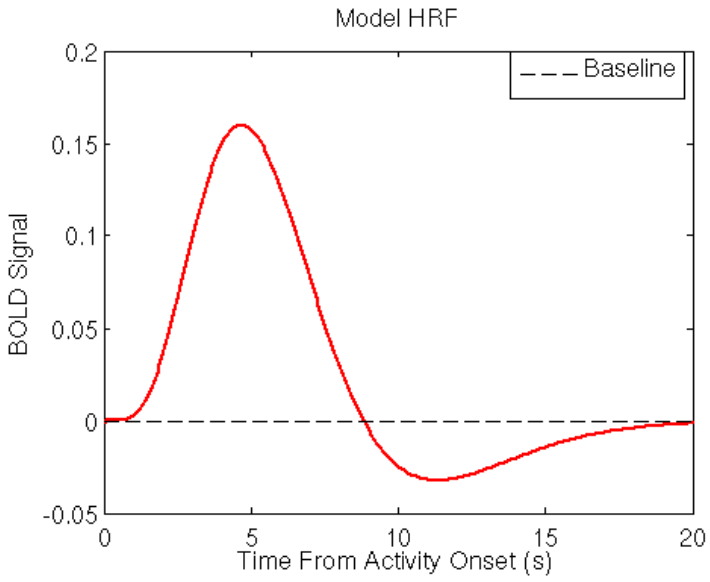


FIGURE 6: A MODEL HAEMODYNAMIC RESPONSE FUNCTION

Because of the delay between stimulus and BOLD peak, fMRI is not well suited to answering questions about when the brain responds to stimuli (in contrast to other methods such as EEG and MEG). However, no other noninvasive neuroimaging method has such a high spatial resolution (single cell recording, which gives the best possible spatial resolution, is highly invasive and can only be used on non-human primates and some epileptic patients undergoing surgery). This means that fMRI is the best tool for discovering what parts of the brain are involved in a given response in healthy humans.

2.3. fMRI SCANNING PROTOCOLS

During a typical fMRI experiment, the participant lies supine in the MRI scanner while performing a task, or responding to stimuli. The RF pulse is applied, and the scanner acquires a volume- a 3D image of the head composed of multiple 2D slices stacked on top of each other. The time required to collect one volume is referred to as the TA, or acquisition time. The time between successive volume acquisitions is called the TR, or repetition time. Acquisition takes place in runs- a block of several volume acquisitions followed by a short break where no scanning occurs before the next run begins. Within each run, data is typically acquired continuously, with no pause between each volume acquisition. However, in some cases- particularly in speech and hearing research- there are also pauses between volume acquisitions.

Because each scan is composed of multiple 2D images that must be closely aligned to reconstruct the 3D volume, it is important for participants to keep their heads still in order to avoid movement artefacts. This is problematic for speech research, since speaking inevitably involves moving the jaw and articulators. In addition, for studies

where it is necessary to present auditory stimuli or record the participant's voice, the noise caused by the switching of the gradient coils introduces significant masking, potentially confounding the experiment. To ameliorate this, a special type of sequence called sparse acquisition is used. In this type of paradigm, single volumes are acquired followed by a pause. Sparse acquisition exploits the fact that the BOLD response does not reach a peak until 5-7 seconds after stimulus onset. This means that it is not necessary to collect data while the task is being performed. Instead, the stimulus is presented in a scanner silent period, and acquisition is timed to coincide with the peak of the BOLD response, after the task has been performed. This has the advantage of reducing interference from scanner noise and head movement during the task, but does result in fewer volume acquisitions, which reduces statistical power. Both of the fMRI studies in this thesis used sparse sequencing.

2.4. fMRI PREPROCESSING

REALIGNMENT

Before fMRI data can be analysed, it is necessary to perform several different preprocessing steps so that participants' data can be compared. In this thesis, preprocessing occurs in the following order: realignment, coregistration of structural with functional images, segmentation of the structural, normalisation and smoothing. The software used was SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), implemented in MATLAB R2013b (Mathworks). An explanation of why each step is necessary and how it is accomplished using SPM8 follows below.

Although movement artefacts can be reduced by using sparse scanning, some head movement in the scanner is still likely. This means that the anatomy that a given set of voxel co-ordinates refer to may change between volumes. This both adds to the residual variance, making it harder to detect true activation (since activation is calculated based on the signal change relative to the residual variance), reduces spatial accuracy, and can result in artefacts- areas of apparent signal difference where in fact there is none. Any changes in signal due to head motion are likely to be much larger than any changes due to brain activity, and can therefore confound the results. In speech studies, where head movement is correlated with the experimental task, these small differences are especially likely to accumulate and result in apparently significant activation. Movement that is time-locked to a task will not be removed by realignment, meaning that it is necessary to use sparse scanning in studies of speech production to mitigate the effects of head movement.

Realigning scans ensures that a single voxel always shows data from the same part of the brain. SPM uses a 6-parameter rigid-body transformation that uses a least squares approach to minimize the difference between the scans and a reference image (usually the first scan in the series). SPM first 'coregisters' the data by calculating the set of movement parameters required to align the source images with the reference. In rigid-body transformation, the reference image remains stationary while the source images are spatially transformed to match it. The type of rigid-body transformation used here is called 'six-parameter' as for a 3-D image, the transformation that maps each voxel in the source image to the same voxel in the reference image needs to specify a translation and a rotation parameter for each of the three dimensions. The least squares method

minimises the sum of squared differences between voxel intensities in the reference and source image. This information is then used to effect a transformation- that is, resample the source images so that they match the reference. If the movement is less than one voxel, interpolation is used to determine the intensity of the transformed voxel when resampling the data. SPM writes out a file containing motion parameters, which can be inspected, and participants excluded if the movement exceeds the minimum dimension of the voxel. In this thesis, data from subjects who moved more than 3mm in any direction over the course of the study were excluded. The remaining participants' motion parameters can then be included in the statistical design as a variable of no interest, allowing nonlinear movement effects to be partialled out of the analysis.

So that activation data can be projected onto an image of the participants' brain, it is necessary to align the T1 structural image to the mean realigned functional image, which is T2 weighted. Because these are two different types of image, it is not possible to realign them using the least squares method to compare voxel intensities; for example, a high voxel intensity in a T1 image might indicate white matter, whereas a voxel with the same intensity in a T2 image would most likely be CSF. Therefore, SPM adopts a mutual information technique, which uses joint probability distributions of intensities in the images to quantify dependencies between the two images- for example, the degree to which low voxel intensity in the T2 image predicts a high voxel intensity in the T1 image.

NORMALISATION AND SEGMENTATION

Once images have been realigned, it is necessary to apply further transformations to participants' brains before they can be compared. Individual brains come in a wide

variety of sizes and shapes, so to make sure that the same brain regions are being compared across subjects, it is necessary to 'normalise' them by warping them them into a common stereotactic space. There are two commonly used stereotactic spaces, or co-ordinate systems. The first is Talairach space, based on the atlas published by Talairach and Szikla in 1967 and updated by Talairach and Tournoux in 1988. Although this co-ordinate system is still used for normalisation by some researchers, it is an imperfect template for warping brains to. The atlas is derived from postmortem dissection of a single human brain, which means that the anatomy is not necessarily typical of the general population. In addition, researchers only dissected one hemisphere of the brain- the left hemisphere is simply a mirrored version of the right. Since human brains are in fact slightly asymmetric (Galaburda, LeMay, Kemper, & Geschwind, 1978), the atlas is not a good representation of a typical human brain in this respect. The second co-ordinate system, used throughout this thesis, is called MNI space, and is an average of many human brains scanned by the Montreal Neurological Institute. There are several different MNI templates based on various numbers of people; the specific template used in this thesis is the ICBM152, which is the average of 152 brains.

The parameters for normalization are generated by segmenting the structural T1 image. This process estimates the probability of tissue types at each voxel and divides the image up into grey matter, white matter, CSF and bone. In the process, the structural image is warped into standard space, and SPM writes out the parameters for this spatial normalization and the inverse normalization. Since the functional images have already been coregistered with the structural, the normalization parameters from the segmentation can be applied to the functional images.

SMOOTHING

Next, the functional images are smoothed. Smoothing helps ameliorate imperfect normalizations resulting from individual variability in subject anatomy. Each voxel is convoluted with a three-dimensional Gaussian kernel, such that the signal intensity at each voxel in the smoothed image becomes the average of the surrounding voxels weighted by the Gaussian. The size of the smoothing kernel is described in terms of its full width at half maximum (FWHM). In this thesis, a smoothing kernel of 8mm FWHM is adopted; this is a medium kernel size, suitable for group analyses.

The effect of smoothing is to remove high-frequency information and 'blur' the image, reducing spatial resolution. Although this may seem counter-productive, smoothing actually increases the signal-to-noise ratio. This is because the noise found in fMRI data is usually centred around zero, randomly distributed, and independent between voxels. Meanwhile, the signal is not randomly distributed, and the signal in one voxel is usually dependent upon its neighbour. Thus, when intensity is averaged across neighbouring voxels, the signal (which is spatially correlated) remains, while the noise will tend to average to zero.

Other possible preprocessing steps include slice timing, which helps account for the fact that each brain volume takes several seconds to acquire, and thus a model based on the first slice of the scan may not accurately account for later slices. Slice timing uses interpolation to model simultaneous acquisition. However, the interpolation is imperfect and can be affected by motion and motion correction. In addition, it is only appropriate when data is acquired continuously, so cannot be used in sparse designs. An alternative is to account for the problem within the statistical model by including a temporal

derivative. This reduces inaccuracies associated with slice timing, but can also reduce the power of the model. Because sparse imaging uses very short acquisition times, the slice timing issue does not hugely affect the accuracy of the statistical model, so it is common not to use a slice timing correction or to model the temporal derivative.

2.5. FUNCTIONAL MRI ANALYSIS: UNIVARIATE

Analysis of the preprocessed functional data takes place in two steps, called first and second level analysis. At the first level, a fixed effect analysis is used to fit the general linear model (GLM) for each subject. Fixed effects analysis estimates the individual variance; it is used to identify significant effects that would be present if the same participants were tested again. Thus, the results tell you only about the specific group that was tested. At the second level, a random effects model is used, which estimates both within-and between-subject variance. The results of the random effects analysis can be generalized to the population that the subjects came from (for example, healthy subjects of the same age).

The GLM models the expected BOLD responses if there were an effect of any condition. Each voxel is analysed individually; this is known as a mass univariate GLM approach. The equation used is:

$$Y = X\beta + e$$

where Y = voxel intensities at each timepoint ('observations'), X= the explanatory variable or variables, β= the beta weights and e= the error term.

The model is specified in SPM using the design matrix. This specifies the onsets and durations of each experimental condition; every row is an observation and every column an explanatory variable. In addition to the conditions of interest, explanatory variables also include motion parameters and other 'nuisance' variables. The onsets and durations of each condition are convolved with a canonical haemodynamic response function (HRF), to account for the delay between neural activity and BOLD response. The canonical HRF is an assumed profile of the vascular response, with the peak set at around 6 seconds followed by a longer overshoot. Once the model has been set up, the next step is to estimate the parameters- that is, to establish which voxels in the data set act as predicted by the design matrix. Estimating the parameters gives an estimate both of effect size and of error for each voxel. This is then written out as a voxel-by-voxel image of the beta weights. Beta weights specify the effect of a given condition at each voxel; if condition A has a greater beta weight than condition B at some voxel, that indicates that A had a greater effect than B. To test for such differences between conditions, a contrast vector can be specified. The resulting contrast image gives a spatial summary of any significant activations resulting from the contrast. T-contrasts are used to compare two conditions and test whether the effect size is greater in one condition than another. F-contrasts compare several conditions, and test whether any of them are different to any of the others. F-tests only specify whether a difference is present, so further statistical tests are required to identify which of the conditions accounts for the difference.

Generally, T-contrasts are created for each condition relative to baseline for each individual subject at the first level, and taken up to the second level for group analysis. At the second level, a new design matrix is constructed with each row representing a single

subject. A one-way t-test with test value of 0 is applied to each voxel to test for a significant effect. This results in an image that provides a map of group effects. Before the results can be interpreted, however, it is necessary to correct for multiple comparisons. Each data set is composed of one test statistic for each voxel in the brain, and the test statistic controls the level of false positive risk only for that voxel- not across the whole image. This means, for example, that if there are 200,000 voxels in the contrast map, at a threshold of $P<0.05$, 10,000 voxels (5% of 200,000) will be false positives due to chance if the null hypothesis is true.

To measure false positive risk over an entire image, then, it is necessary to correct the p-threshold to account for the number of tests that are being performed. There are two common types of correction for multiple comparisons- familywise error (FWE), and FDR (false discovery rate). The FWE rate is the chance of one or more false positives anywhere in the image. A corrected FWE p-value of $P<0.05$ means that there is at most a 5% chance of false positives in the thresholded map. The simplest type of FWE correction is the Bonferroni correction, which divides the p-threshold by the number of tests. However, a requirement of the Bonferroni correction is that each statistical test be independent. This is transparently not true of neuroimaging data, where activation at a given voxel is often correlated with those around it. This means that although the Bonferroni method adequately corrects Type I error (false positives), it also results in a very high rate of Type II error (false negatives). A more commonly used FWE correction is random field theory, which attempts to account for spatial correlation in the data by assuming that it mimics a smoothly varying random field. However, this is still a very conservative correction (Nichols & Hayasaka, 2003).

Since fMRI data is inherently noisy, even after preprocessing it is almost inevitable that some voxels will appear active just by chance. This means that a FWE correction at $p<0.05$ is attempting to establish something that is very unlikely for this type of data- a 5% chance that no voxel is a false positive. FDR correction is an alternative approach, which assesses the probability that (at FDR $p<0.05$) no more than 5% of voxels are false positives. The advantages of FDR correction are increased sensitivity and flexibility: as the correction is based on the distribution of p-values, it takes into account the amount of signal in the data and is thus more sensitive with a small signal, and more conservative with a large signal. The drawback is an increased number of false positives compared to FWE correction. However, false positives are likely to be randomly spread throughout the data, while meaningful voxels are more likely to be clustered together, reflecting populations of active neurons. This means that a small amount of noise is generally not problematic, as researchers are more interested in patterns of activation than individual voxels. In this thesis a combination of FWE and FDR correction approaches are used.

2.6. Functional analysis: Regions of interest

A region of interest (ROI) analysis restricts the analysis only to a particular area, usually one in which there is an a priori hypothesis about the area's involvement in the task. By limiting the search volume, ROI analysis reduces the problem of multiple comparison, meaning that it is more sensitive to changes in the signal. In addition, it can give a better idea of what is happening over an anatomical region than looking at the timecourse of an individual voxel within that region, which might be an outlier. ROIs can be defined based on anatomical criteria, co-ordinates of interest from previous studies, or from the results

of the random effects (RFX) analysis. However, when defining an ROI based on RFX results, it is important to choose selection criteria carefully to avoid circular analysis. Circular analysis, also known as non-independence or 'double dipping', occurs when the same contrast or related contrasts are used to select an ROI and to analyse the data within it. For example, in an experiment with three conditions, A, B and C, if the contrast A>B is used to define an ROI, that ROI cannot then be used to test the hypothesis that condition A has a greater effect than condition B or C. By virtue of the selection criteria, we already know that voxels in the ROI are more likely to be active in condition A. Therefore, any analysis of the ROI that looks for more activation in condition A than condition B or C will exaggerate the effect (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

2.7. FUNCTIONAL ANALYSIS: INDEPENDENT COMPONENT ANALYSIS

Univariate analysis techniques identify whether, at each individual voxel in the brain, the average BOLD response to one condition is greater than to another. There are two potential disadvantages of this approach. First, it assumes that all voxels are independent of one another, and this fails to account for the fact that processes in the brain are usually organized into distributed networks (Fox et al., 2005). Second, with this approach it is only possible to study activation that has been modelled beforehand, meaning that the analysis is dependent on assumptions about the timecourse of the data. Independent Component Analysis (ICA) is a blind source separation method, which attempts to segregate the observed signal into its hidden sources (components). These sources are assumed to be statistically independent and non-Gaussian with an unknown linear mixing process. Because ICA requires no explicit temporal model,

results are data-driven rather than dependent on prior assumptions. It is also better at addressing the correlated nature of neural data than other models such as Principal Component Analysis (PCA); where PCA looks at relationships between pairs of voxels and is thus limited in what inferences it can draw about whole networks, ICA is a hugely multivariate approach that considers the correlations between all voxels simultaneously.

ICA can be used to discover either spatially or temporally independent components. This thesis used spatial ICA (sICA) as implemented by GIFTv3.0a (http://mialab.mrn.org/software/gift/) to reveal maps of brain networks that each explain unique variance of the time series. These components are maximally independent spatially, but may overlap in time, meaning that the analysis can be used to extract components associated with overlapping events such as hearing noise at the same time as speaking over it.

Once components have been extracted, it is necessary to separate them into components that represent fluctuations in the signal associated with functional brain activation, and 'noise' components that are not of interest. Noise may result from physiological factors such as activation in areas such as ventricles and venous sinuses that are affected by breathing and heart rate, or from other factors such as scanner drift or motion artefact. In this thesis, a combination of automated and manual processing was used to identify artefactual components. The different component maps were spatially correlated with maps of white matter and CSF as well as with different maps of probabilistic networks provided in the GIFT toolbox. Components that had a high

correlation with non-grey matter tissues were excluded. Signal-related components usually consist of a low number of large clusters in grey matter, that follow known anatomical boundaries, display a regular oscillatory time course and are predominantly low frequency (Griffanti et al., 2016); any components that did not meet these criteria were likewise excluded. Finally, differences between experimental conditions can be revealed by correlating each component with information about condition onsets and durations, as entered into the SPM design matrix. Resulting contrast maps can be thresholded and viewed in SPM in the same way as the contrast map from a univariate analysis.

2.8. fMRI META-ANALYSIS: ACTIVATION LIKELIHOOD ESTIMATION

Because of the expense and difficulty of carrying out fMRI studies, sample sizes are often low, making it difficult to draw conclusions from the results of any one study. Meta-analysis techniques address this by aggregating information from multiple studies to find consistent responses. Activation Likelihood Estimation (ALE) is a type of meta-analysis that assesses convergence between fMRI co-ordinates reported in different studies. ALE treats activation foci as spatial probability distributions centred at given co-ordinates. The analysis asks if there is any convergence among foci that cannot be explained by a hypothetical null distribution in which the foci are randomly distributed throughout the brain. Specifically, it is concerned with convergence between (rather than within) studies. To assess this, all foci from the same study are modelled as Gaussian probability distributions; the size of the Gaussian kernel at FWHM is determined by a model of spatial uncertainty. The distributions are then merged to create a single 3D volume or 'modelled

activation' (MA) map for that study. Each subject's MA map is entered into the analysis and used to calculate an ALE statistic for each voxel in the brain from the union of the MA probabilities. The ALE statistic gives the probability of activation being present at that voxel for all studies in the analysis.

Any meta-analysis is affected by the quality of the studies it contains, so it is important to apply consistent guidelines to ensure that the analysis is both robust and replicable. In order to identify data for a meta-analysis it is usually necessary to first conduct a systematic review of available studies. In this thesis the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, or PRISMA statement (Moher et al., 2009) was used. The PRISMA statement consists of a flow diagram and checklist, intended to guide the process of selecting studies and reporting the results transparently. In brief, a systematic review involves a comprehensive search of the literature with a set of eligibility criteria in mind. The eligibility criteria may be constructed based on the acronym PICOS: Population, Interventions, Comparator, Outcomes, Study Design. In other words, the criteria typically specify that to be included, a study must have tested a particular population, used a specific method or intervention, included an appropriate control group, and reported the same type of outcome- for ALE studies, this would include peak voxel co-ordinates gained from whole brain analysis. The number of studies that met the criteria as well as those that were rejected are reported. The eligibility criteria must be stated, as well as the databases that were searched and the keywords used. Once a thorough search has been made, the results of the studies found are synthesised and presented in review form. This involves critically appraising the validity of the included studies, for example through assessment for bias. Bias is possible both in the study design

(for example if participants were not naive to the intended outcome) and at outcome level (for example if the measure reported is subjective or unreliable). Studies at a high risk of bias may need to be excluded from the meta-analysis.

In this thesis, a systematic review was conducted to assess the strength of the evidence for an effect of manipulating auditory feedback on neural activation in the superior temporal gyrus. The co-ordinates reported in these feedback studies were then used to conduct an ALE looking for convergence in the reported foci. The results are reported in the following chapter.

CHAPTER 3: A SYSTEMATIC REVIEW AND ALE META-ANALYSIS OF ALTERED AUDITORY FEEDBACK STUDIES USING FMRI AND PET.

3.1. ABSTRACT

Evidence for different models of speech production is often drawn from investigations in which the sound of a talker's voice is altered in real time. Methods of feedback manipulation vary, but are assumed to engage the same neural network and psychological processes. This review aimed to compare behavioural and neural outcomes of different of feedback alteration techniques and assess the strength of the evidence for models of speech production. A systematic review of articles written in English was conducted using PubMed and Web of Science. Search terms included 'speech', 'auditory feedback', 'fMRI' and 'PET'. Only functional neuroimaging studies of speech production carried out using fMRI or PET, in neurotypical adult humans, using one of three predefined auditory feedback techniques (frequency altered feedback, delayed auditory feedback and speech in noise) were included. Extraction of data from articles was carried out by a single author, using predefined data fields based on the Cochrane handbook. The co-ordinates of brain areas that responded preferentially to altered feedback over unaltered feedback were analysed using the GingerALE toolbox (www.brainmap.org). These foci predominantly clustered in superior temporal gyrus. To sum up, a superior temporal gyrus response appeared across all studies regardless of type of feedback used. However, this was not always statistically robust or correlated with expected behavioural changes.

3.2. INTRODUCTION

RATIONALE

The two speech production models discussed in Chapter 1 both suggest that error correction during speech production is achieved by a feedback circuit that compares auditory feedback to an internal target or model, then issues corrective signals. Both models hypothesise that this feedback monitoring and error correction takes place in the superior temporal gyrus (STG).

Since error production and correction in natural speech is unpredictable and sporadic, researchers wishing to investigate the mechanisms of speech error correction have relied on various methods of introducing external 'errors'. This review looks at three techniques commonly used to alter speech feedback: frequency shifted feedback (FAF), delayed auditory feedback (DAF) and masked auditory feedback (MAF). Though they differ in the type of perturbation used and the assumed auditory target, all three manipulations are presumed to prompt the same error correction mechanism. This chapter reviews the evidence for a common error correction mechanism and for the role of the STG in feedback control. Since the STG is a functionally heterogenous area, as a follow-up an ALE meta-analysis looks for convergence in reported co-ordinates between studies, in an attempt to pinpoint the region of STG involved in error correction.

OBJECTIVES

To investigate the role of the superior temporal gyrus in feedback control, we reviewed functional imaging studies that used one or more of the three specified feedback alteration techniques (FAF, DAF or MAF). To be characterised as a feedback response it

is necessary for the intervention to prompt both a neural reaction and behavioural compensation, but often this is a secondary consideration in functional neuroimaging. Therefore, the strength of the studies was assessed both on whether reported neural activations were robust enough to survive correction for multiple comparison, and whether behavioural results supported the interpretation of neural data. Additionally the ecological validity of tasks used is considered. Neural co-ordinates resulting from a univariate comparison of brain activation when speaking with altered feedback versus speaking with normal feedback were entered into an ALE meta-analysis.

3.3. METHODS

Searches using the keywords 'speech', 'auditory feedback', 'fMRI', 'magnetic resonance imaging', 'PET', and 'positron emission tomography' were used to identify studies for inclusion, using the electronic databases PubMed and Web of Science. The search was conducted in February 2016 and yielded 144 results. 42 duplicates were removed and then the remaining 102 studies were assessed for inclusion based on their abstracts. Those records selected for inclusion were studies of speech production in humans, published in English, that used one of the three specified altered feedback techniques in combination with a functional neuroimaging method (fMRI or PET). 87 studies that did not meet all of these criteria were excluded at this stage. Where it was unclear if a study met the inclusion criteria based on its abstract, the full text was read and assessed. Figure

7 shows a breakdown of the study selection process.

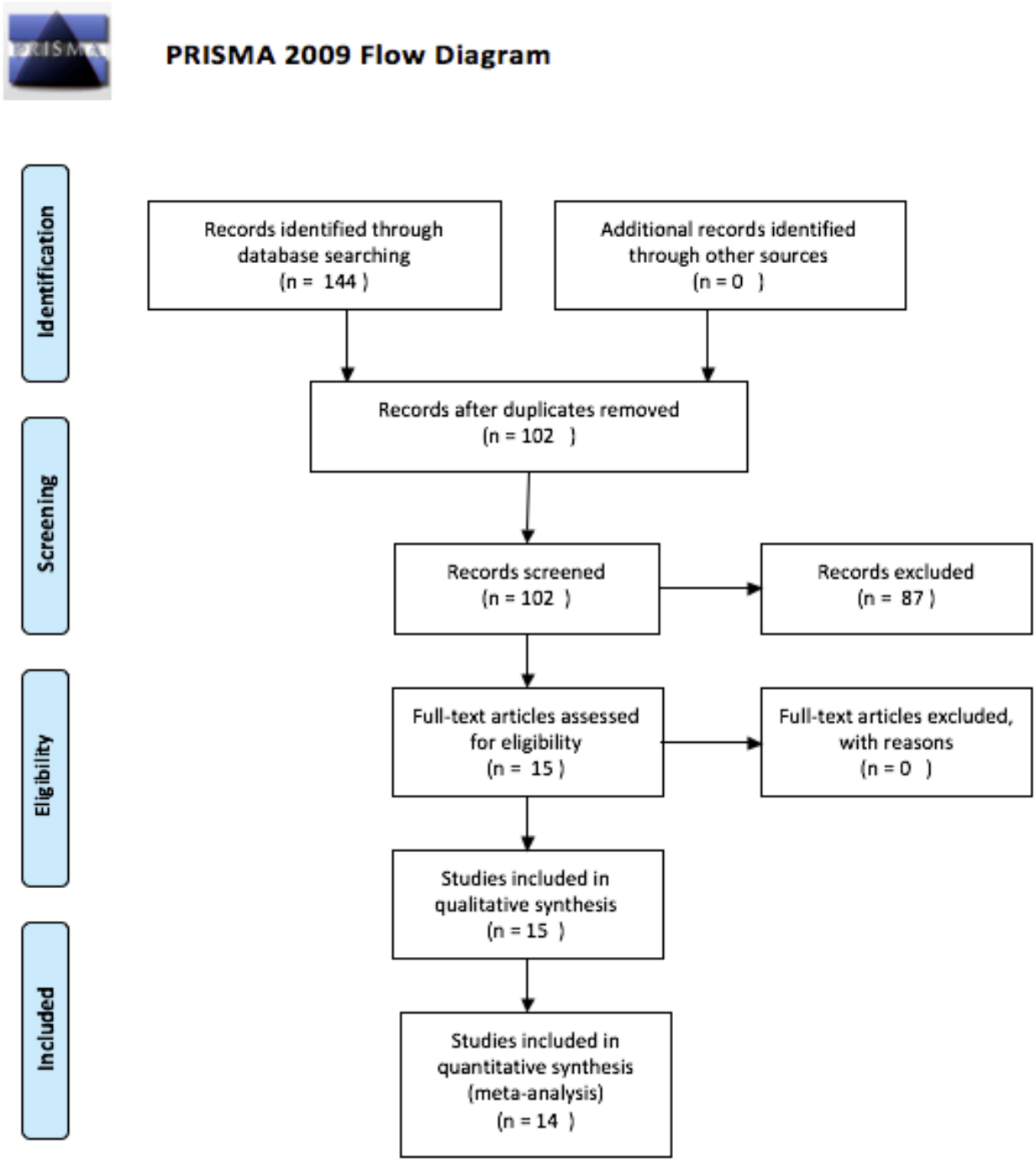


FIGURE 7: PRISMA FLOW DIAGRAM OUTLINING STUDY SELECTION PROCESS

DATA COLLECTION & DATA ITEMS

Information was extracted from each included trial on (1) Participants (number, inclusion criteria, age and gender); (2) Task performed (feedback alteration type, speech production task, other experimental and control conditions); (3) Neural data acquisition and analysis (acquisition parameters, stereotactic space, corrections for multiple comparisons, region of interest analyses, any other statistical methods used); (4) Behavioural data and results (measures of vocal compensation); and (5) Neural results for the altered vs unaltered feedback comparison. The full data extraction sheet is included in Appendix A.

RISK OF BIAS

One factor that can affect the reliability of neural results is the choice of significance threshold and correction for multiple comparisons. Eklund, Nichols and Knutsson (2016) recently found that using a clusterwise threshold of FWE-corrected $p<0.05$ results in false positives far in excess of the expected 5% (up to 50% false positives depending on the analysis software used), while a voxelwise threshold of FWE $p<0.05$ is frequently too conservative. In this analysis, studies were judged less reliable if they did not correct for multiple comparisons at the peak level- this includes studies using uncorrected statistics as well as those corrected at cluster level. On the other hand, studies that found no results at a voxelwise threshold of FWE $p<0.05$ may not have found an effect because of the FWE correction's relative lack of sensitivity. In addition, studies were judged on their inclusion of an appropriate control condition. Feedback-sensitive areas can be identified by the presence of speaking-induced-suppression (SIS)- a reduced response to vocalising compared to hearing similar sounds. Studies that did not include a listening control were

therefore less able to demonstrate that neural activations were associated with a feedback response. Finally, since the neural mechanism under investigation is purportedly an error correction system, any neural activation should be associated with a behavioural correction or compensation to the 'error', or perturbation. Again, studies that did not analyse behavioural results, or found no behavioural compensation, were considered less reliable evidence than those that did.

SUMMARY MEASURES

There were two outcomes of primary interest. Brain co-ordinates that showed a preferential response for altered feedback were collected for the ALE analysis. Typically these co-ordinates were the result of a contrast between speaking with altered feedback and speaking with no alteration. Some studies used more nuanced methods to identify feedback-sensitive brain regions, for example by parametrically varying the perturbation and looking for an associated brain response, or by factoring out responses to listening to speech as well as producing it. Co-ordinates resulting from such analyses were also included in the meta-analysis.

The second outcome of interest was evidence of behavioural adaptation to altered feedback- for example, changes in pitch to oppose a frequency shift, an increase in intensity to compensate for noise masking, or a change in speech rate in response to delayed auditory feedback. To be able to draw a strong conclusion about brain regions involved in feedback control, it is essential that neural activation be accompanied by behavioural evidence that feedback control is taking place.

PLANNED METHODS OF ANALYSIS

Peak voxel co-ordinates from areas identified as sensitive to altered feedback were collected from each study. Co-ordinates in Talairach space were converted to MNI space space using the Brett transform (Brett, Christoff, Cusack, & Lancaster, 2001) and submitted for analysis using the GingerALE software (www.brainmap.org). Results were corrected for multiple comparisons using a voxel-level threshold FWE $p<0.05$, recommended by Eickhoff (Eickhoff et al., 2016) as the least likely to result in false positives with a sample size of fewer than 17 studies.

TABLE 1: STUDIES INCLUDED IN REVIEW

Source	Number of subjects	Vocalization task	Feedback alteration	Population description	Imaging method	Threshold	Contrast	Number of foci at whole brain level
McGuire, Silbersweig, & Frith, 1996	6	Words	FAF	Non-clinical	PET	Uncorrected p<0.001	Pitch shift>normal feedback	4
							Substitute voice>normal feedback	5
Hirano et al., 1997	6	Sentences	DAF	Non-clinical	PET	Uncorrected	Delay>rest	10
Hashimoto & Sakai, 2003	18	Sentences	DAF	Non-clinical	fMRI	P<0.05, corrected (correction not specified)	Delay>normal feedback	6
Fu et al., 2006	13	Words	FAF	Non-clinical	fMRI	Cluster level p<0.01	Pitch shift>normal feedback	21

Christoffels, Formisano, & Schiller, 2007	14	Words	MASK	Non-clinical	fMRI	Cluster level p<0.05	Noise>normal feedback	3
Toyomura et al., 2007	12	Phoneme	FAF	Non-clinical	fMRI	Corrected p<0.05	Pitch shift>normal feedback	5
Tourville, Reilly, & Guenther, 2008	10	Words	FAF	Non-clinical	fMRI	FDR p<0.05	Pitch shift>normal feedback	None
Takaso, Eisner, Wise, & Scott, 2010	8	Sentences	DAF	Non-clinical	PET	Uncorrected p<0.0001	Parametric response to delay increase	5
Zarate, Wood, & Zatorre, 2010	9	Phoneme	FAF	Singing	fMRI	FWE p<0.05	Conjunction of two pitch shift conditions>normal feedback	15
Zheng, Munhall, & Johnsrude, 2010	21	Word	MASK	Non-clinical	fMRI	P<0.05, corrected (correction not specified)	Interaction between masked>normal feedback and listening>normal feedback	2

Christoffels, van de Ven, Waldorp, Formisano, & Schiller, 2011	11	Word	MASK	Non-clinical	fMRI	FWE p<0.05	Parametric response to noise masking	None
Parkinson et al., 2012	12	Phoneme	FAF	Non-clinical	fMRI	Uncorrected p<0.001	Pitch shift>normal feedback	6
Niziolek & Guenther, 2013	15	Word	FAF	Non-clinical	fMRI	FDR p<0.05	Pitch shift>normal feedback	8
Zheng et al., 2013	16	Word	FAF/MASK	Non-clinical	fMRI	FWE p<0.05	MVPA (pitch shift & mask) > (pitch shift & no-shift) AND (mask & no-shift	4
Behroozmand et al., 2015	8	Phoneme	FAF	Epileptic	fMRI	FWE p<0.05	Pitch shift>normal feedback	None

3.4. SUMMARY OF STUDIES INCLUDED

Fifteen studies were included in the review. Apart from Behroozmand et al. (2015), who used patients awaiting surgery for epilepsy, and Zarate, Wood, & Zatorre (2010), who recruited singers, all subjects were neurotypical right-handed men and women with no hearing or language impairment, or musical expertise. Although Zarate et al. (2010) explicitly recruited musicians and used phoneme production as a singing task, it was considered that the study was similar enough to other FAF phoneme production studies that it merited inclusion. In total there were 180 participants across all 15 studies (107 males; 73 females), aged between 18 and 53 years old (mean age 28). Three studies used PET imaging while the rest used fMRI.

Eight out of fifteen studies were frequency altered feedback (FAF) studies; three used delayed auditory feedback (DAF), three used noise masking (MAF) and one used both frequency altered feedback and noise masking. Eight studies asked participants to produce single words, either by reading aloud or by naming pictures; four studies used phoneme vocalisation, and three used whole sentences as a vocalisation task. The number of different stimuli (words, phonemes or sentences) used in each experiment varied from 360 (McGuire et al., 1996) to just one (Zheng et al., 2010) The exact choice of task and stimulus is discussed below.

PERTURBATION TYPE

Although the type of perturbation can broadly be divided into DAF, FAF and MAF, there was considerable variation in the exact manipulation that each study used. For example, frequency altered feedback can vary in the degree of perturbation (that is, how many

cents or semitones the talker's utterance is shifted by), the direction of perturbation (up or down), and the part of the utterance that the shift was applied to. Of nine studies that used used a frequency altered feedback paradigm, only two studies (Toyomura et al., 2007; Zarate et al., 2010) used the same frequency manipulation (a pitch shift of 200 cents), and even so there were several other differences between the studies, both in stimulus and in task design. Across the nine studies, shifts ranged from 25 cents (0.25 of a semitone) to 8 semitones (800 cents) in size. The direction of the pitch shift also varied between studies, with some (Tourville et al., 2008; Parkinson et al., 2012; Toyomura et al, 2007) using both up and down pitch shifts and others using just one or the other.

While most studies shifted the frequency of the whole utterance, three studies shifted just the first or second formant (F1 and F2). In vowels, F1 is inversely related to height (the lower the first formant, the higher the vowel) while F2 affects vowel backness (a higher F2 means a more fronted vowel). Thus, changing F1 and F2 can affect which vowel participants heard, producing an 'error' at the phoneme level. Zheng et al. (2013) altered the first and second formants such that the participant vocalised 'head' but heard a vowel that was closer to 'had'. Niziolek et al. (2013) used a subjective paradigm in which F1 and F2 were shifted differently for each subject depending on where their phoneme boundary lay, such that in one condition the shift remained within phoneme boundaries ('hid' remained 'hid'), while in the second condition, the shift crossed the boundary ('hid' became 'had'). Tourville et al. (2008) altered just the first formant, shifting it either up or down by 30%. The remaining studies applied frequency alterations to the whole utterance.

Three studies used delayed auditory feedback: Hashimoto & Sakai (2003), Takaso et al (2010) and Hirano et al (1997). Responses to DAF depend on the choice of latency- the time between speaking and hearing the feedback. Research (Stuart, Kalinowski, Rastatter, & Lynch, 2002) suggests that a delay of 200ms is maximally disruptive; accordingly Hashimoto and Sakai (2003) chose a latency of 200ms, while Takaso et al. (2010) chose to vary latency to investigate regions associated with increasing disfluency, and used three latencies: 50, 125 and 200ms. However, Hirano et al. (1997) used a latency of 100ms; the reason for choosing this particular latency is unclear.

Finally, four studies used masked auditory feedback. Here, the potential for variation is much greater, since technically any sound can be a masker. Additionally, it is important to choose a stimulus intensity at which speech is effectively masked, without causing hearing damage. Zheng et al. (2010; 2013) used signal-correlated noise (white noise modulated with the amplitude envelope of speech) played at 85dB. Christoffels et al. (2007; 2011), meanwhile, used continuous pink noise. Pink noise is noise with equal intensity in each octave (so, compared to white noise, it has more intensity at low frequencies). The noise was set at a level subjectively judged loud enough to prevent the participant from hearing their own voice; consequently, the maximum noise level was different for each participant. In Christoffels et al. (2011), this maximum intensity was then decreased by 10dB SPL and 15dB SPL respectively to create two further noise conditions.

OTHER EXPERIMENTAL CONDITIONS

In addition to the manipulation created by the three different types of feedback alteration, some studies introduced additional manipulations. While in most FAF studies, the manipulation was covert and participants were not explicitly instructed on how to respond, Zarate et al. (2010) asked participants, who were trained singers, to consciously ignore or compensate for the pitch shift. Fu et al. (2006) and McGuire et al. (1996) both included a condition where the subject heard their own voice replaced (overdubbed) by another's. Additionally, Fu et al. (2006) also shifted the frequency of this voice, such that when participants spoke, they heard their own voice with and without frequency shift, and a substitute voice with and without frequency shift. Hashimoto and Sakai (2003) asked participants to speak at their normal speaking rate, rapidly, and slowly. This was intended to control for any neural activation shown when speaking under delayed auditory feedback that could be attributed to a change in speech rate. Hirano et al. (1997) also included a condition in which the participants' voice was low-pass filtered.

Seven studies also included a listening control condition in which participants heard one of the masking sounds, or their own voice either with or without the manipulations. However, more than half of the studies did not include a listening control.

SPEECH PRODUCTION TASK

The nature of the speech production tasks used is important because it is possible that different levels of articulation are processed differently. For example, in speech production, the Hierarchical State Feedback Model (Hickok, 2014a) posits that speech is controlled at the phoneme level by a motor feedback loop and at the syllable level by

auditory feedback. However, the choice of task is necessarily the result of an interplay between sparse fMRI scanning constraints (which require the utterance to be as short as possible) and feedback alteration type. Adaptation to frequency altered is feedback strongest when perturbation duration exceeds 100ms (Burnett, Freedland, Larson, & Hain, 1998) so requires extended vocalisation to prompt behavioural adaptation. Consequently, four out of nine FAF studies used single phoneme vocalisation. These four studies all used just one stimulus- the phoneme /a/- and required talkers to produce the phoneme continuously for four (Zarate 2010) or five seconds (Behroozmand et al., 2015; Parkinson et al., 2012; Toyomura et al., 2007). Some other studies that used single words also required talkers to prolong their utterances (Niziolek et al., 2013). A typical articulation rate in conversational speech is 10 phonemes per second (Osser & Peng, 1964). Producing one phoneme over five seconds is therefore fifty times slower than typical articulation; closer to singing, in that it requires extended breath control— and in fact Zarate et al. (2010) used it as a singing task.

For studies that used formant manipulation, it was necessary to ensure that each stimulus contained a vowel that could be manipulated, so Niziolek et al. (2013), Tourville et al. (2008), and Zheng et al. (2010; 2013) all used lists of monosyllabic consonant-vowel-consonant (CVC) words, such as 'bed'. These lists were very short: Tourville et al. (2008) and Niziolek et al. (2013) used eight words containing /ɛ/, while Zheng et al. (2013) used just two: 'had' and 'head', and Zheng et al. (2013) had just one stimulus word, 'Ted'. The two remaining FAF studies, Fu et al (2006) and McGuire et al. (1996), manipulated the frequency of whole words, and perhaps as a result used a much wider variety of stimuli, asking participants to read from lists of 96 and 360 words respectively.

All of the DAF studies used whole sentences, as delayed auditory feedback is effective over long periods of connected speech (Yates, 1963). Additionally, two out of the three DAF studies used PET, in which continuous scanning during speech is possible with no scanner noise, so there are fewer constraints on task length. It should be noted, however, that it is possible to use whole sentences (albeit short ones) as stimuli in an fMRI paradigm without too much difficulty, as demonstrated by the fact that the third DAF study, which used 17 seven-syllable sentences, was an fMRI study (Hashimoto and Sakai, 2003). It is therefore surprising that all the speech production in noise studies, which might also be expected to use sentences to maximise the possibility of behavioural adaption, instead used single words- and in fact took steps to avoid compensation for masking noise. Christoffels et al. (2007; 2011) instructed participants not to raise the level of their voice in response to the noise, while Zheng et al. (2013) asked subjects to whisper.

3.5 Behavioural adaptation and associated neural responses

Acoustic analysis

Acoustic data can help to interpret the neural response by showing if adaptation has taken place. If a neural response is associated with feedback control then it should be followed by a vocal correction for the 'error'.

Six studies, of which five were FAF studies and one was a DAF study, reported the results of acoustic analysis. FAF studies typically reported response direction, magnitude and latency, although the exact calculations used varied between studies. Overall, participants tended to shift their voices in the opposite direction to the manipulation ('opposing' the shift), within 200ms of stimulus onset- a very slow response in comparison to typical speech rates (Osser & Peng, 1964). However, responses were often inconsistent on a trial-by-trial basis, and response magnitude was typically only a small fraction of the size of the perturbation. Parkinson et al. (2012) defined response magnitude as the difference between the baseline mean F0 and the point of greatest F0 deviation during the shifted trials, while latency was measured as the time between stimulus onset and the response exceeding two standard deviations of the baseline mean. All participants had previously passed a pre-screening that tested for the feedback adaptation response, and they also demonstrated vocal adaptation in the experiment itself. In response to a frequency perturbation of 100 cents, they shifted their voice up by 21.36 cents (s.d. 10.88) on average in response to downward shifted stimuli, and down by 17.47 cents (s.d. 5.48) for upward shifted stimuli. Mean response latency was 232ms for the downwards shift and 202ms for the upwards shift. Behroozmand et al. (2015) also

found that participants opposed the frequency shift, with a mean vocal response magnitude of 4.5 cents. This was significantly different to the vocal response magnitude at baseline, although less than 1% of the experimental pitch perturbation (of 600 cents).

Niziolek et al. (2013) measured response magnitude as a percentage of the shift magnitude. Latency was defined as the time at which the response magnitude significantly differed from baseline for five time points in a row. The trials were divided into shifts that fell 'Near' and 'Far' from the phoneme category boundary in a post-hoc analysis, as the predefined 'Across' category boundary shifts did not always result in a phoneme boundary being crossed due to speaker variability. Response magnitude was greater in trials that were closer to the phoneme boundary (25% response magnitude) than trials that were far from the boundary (response magnitude 3%). Latency was shorter for Near trials (140ms) than for Far trials (256ms). As participants did not consistently change their voices in the opposite direction to the shift, Niziolek et al. (2013) also measured the efficiency of compensation, defined as the percentage of total vocal deviation that opposed the shift. This was also greater in the Near condition than in Far trials.

Tourville et al. (2008) found that mean compensation was 30Hz (or 13.6% of perturbation magnitude) in the shift up condition and 28.3Hz (13% of perturbation magnitude) in the shift down condition; there was no significant difference between shift directions. Zarate et al. (2010) found that when asked to ignore pitch shifts, singers were less able to suppress responses to the 25 cent pitch shifts than to the 200 cent shift. When consciously compensating for the shift, they under-compensated for the 200 cent pitch

shift (compensation 87.66% of perturbation magnitude) and over-compensated for the 25 cent pitch shift (compensation 112.67% of perturbation magnitude).

Hashimoto and Sakai (2003) used a delay index measuring the number of correctly spoken morae per second (a rhythmic unit in Japanese speech, similar to a syllable), compared to speech at baseline. This showed that participants were less fluent in the delayed condition.

OTHER BEHAVIOURAL ANALYSIS

Takaso et al. (2010) did not collect acoustic data, but asked participants to rate their speech for speed, perceived difficulty, and accuracy of articulation, on a scale from 1 to 10. They found that as the delay increased, so did perceived difficulty, while speed and accuracy ratings decreased.

Fu et al. (2006) and McGuire et al. (1996) who both used a substitute voice as part of their manipulation, asked participants to attribute the voice they heard to 'self', 'other', or 'unsure'. McGuire et al. (1996) found no misattribution of feedback, but did not ask participants on a trial-by-trial basis. Fu et al. (2006) by contrast, found that altered feedback led to considerable misattributions and unsure responses, but that participants were able to correctly identify their voice and a strangers voice, although there were more misattributions in the substitute voice condition. Participants made significantly more attribution errors when speaking with FAF condition than when speaking without feedback alteration, but there was no difference in errors between substitute voice conditions (with and without pitch shift). Participants were also more likely to make

'unsure' responses in the presence of a pitch shift regardless of the voice's source (self or other).

Six studies did not report behavioural adaptation data. Hirano et al. (1997) reported that talkers spoke fluently in the low-pass filtered condition and dysfluently under delayed auditory feedback, but did not present any behavioural data to support this. Toyomura et al. (2007) did not report data for all participants, but provided a sample figure showing changes in pitch trajectory under normal and pitch-shifted feedback; this sample appears to show compensation. McGuire et al. (1996) commented anecdotally that speakers tended to shift their speech during the pitch distortion task such that their voice also sounded distorted. The four studies that used masking noise instructed participants not to make any adaptation response, and Christoffels et al. (2007; 2011) reported that participants successfully suppressed the Lombard response. Zheng et al. (2010; 2013) also argued that because of the study design (in which conditions were randomised on a trial-by-trial basis rather than presented in blocks), they could not measure behavioural data- although, as other studies found, behavioural adaptation can occur within 200ms of perturbation onset, and each trial was 1.6s long.

3.6. FMRI CHOICE OF COMPARISON AND THRESHOLD

The exact choice of contrast is important in identifying neural activation as a feedback response. A canonical feedback response is one in which there is comparative suppression when speaking without altered feedback compared to listening, and speaking in altered feedback results in a release from this suppression. Few studies made this three-way comparison, however; indeed, eight out of the fifteen studies were unable

to as they did not include a listening control condition. Of those that did, Behroozmand et al. (2015) found no regions in which activation was greater when listening to speech recordings compared to speaking with normal feedback, although the reverse comparison (speaking>listening) did elicit activation in motor and somatosensory cortex. Additionally, no activation was seen for the shift>no shift speech production conditions except at an uncorrected threshold of $p<0.001$. Christoffels et al. (2007) similarly found no significant difference between the average BOLD response to speaking in noise compared to listening (either to the noise stimuli or to recordings of the participant's voice). However, there was significant activation in the STG when speaking with noise was contrasted with speaking without noise. In a follow-up study designed to investigate whether increased masking levels resulted in greater STG activation (Christoffels et al., 2011), they found no significant effect of masking intensity at the whole-brain level. However, when the analysis was masked using an auditory localizer, they found that auditory cortex activity decreased during unmasked feedback, but increased parametrically in line with masker intensity when participants spoke in noise. An F-test confirmed that this parametric neural response was seen only when participants spoke in noise and not when they listened to the same stimuli. McGuire et al. (1996) found that speaking with normal feedback was associated with decreases in frontal and parietal activation when compared with listening to speech, but did not find a speaking-induced suppression response in STG. The two altered feedback conditions- pitch shifted feedback and overdubbed speech- were contrasted with reading aloud with unaltered feedback. Both contrasts resulted in increases in activation in bilateral temporal cortices. Zarate et al. (2010) did not report a comparison of regions that were more active for

listening than for vocalising, but did make the reverse comparison, which found that bilateral auditory cortex was more active when participants spoke than when they listened. Specifically, singers recruited bilateral planum temporale and right BA6/44 when ignoring pitch shift. When consciously compensating for the shift they recruited a network including right STS, planum temporale, motor and somatosensory cortex.

Zheng et al. (2009) used an F-test to look for an interaction between condition and task type, such that speaking in masking noise yielded more activation than speaking without a masker or listening to playback of their own voice. This resulted in bilateral posterior STG activation. Zheng et al. (2013) used multivariate pattern analysis to look for networks where there was a difference between altered and unaltered feedback in perception but not production. This included bilateral STG/MTG and left pre central gyrus.

Most other studies compared the altered feedback condition with speaking with unaltered feedback. There are some problems with using this contrast to identify feedback-sensitive regions- all of the feedback manipulations involved talkers hearing something other than their own voice, and since the STG is active in audition it may be that enhanced STG responses to altered feedback are simply a response to hearing something unusual. With this caveat in mind, most studies comparing altered feedback with normal feedback found the expected response in STG. A list of all significant co-ordinates found in each study and their probabilistic anatomical correlates is shown in Table 2 below. Hashimoto & Sakai (2003) contrasted DAF with normal feedback and found activation in STG and supramarginal gyrus. Additionally, behavioural measures of speech fluency were positively correlated with activation in bilateral STG. Niziolek et al.

(2013) used a correlation analysis which showed that behavioural adaptation was positively correlated with activation in the STG and inferior frontal gyrus. Toyomura et al. (2007) contrasted FAF with no FAF to find activation in several regions including the STG and insula. Fu et al (2006) also compared FAF with no FAF at a cluster level significance threshold of $P<0.01$ and found activation in STG, IFG and anterior cingulate cortex.

Some studies failed to find activation at whole brain level. Parkinson et al. (2012) found activity in STG for the FAF>no FAF contrast, but only at the very low threshold of uncorrected $p<0.001$. Tourville et al. (2008) also failed to find activation at the whole brain level in a random effects analysis. A fixed effect analysis showed activation in bilateral posterior STG and planum temporale; they then conducted ROI analysis on 142 regions, identified as potential feedback regions by the DIVA model. This also revealed activation in posterior STG, although as the ROI analysis was only carried out after the fixed effect analysis confirmed where activation could be found, it is possible that this is circular analysis. Finally, Hirano et al. (1997), in addition to not correcting for multiple comparisons, did not contrast delayed with normal feedback, but with rest, which makes the results difficult to interpret. However, activation was observed in bilateral STG as well as IFG, motor and visual cortices and cerebellum. DAF contrasted with low-pass filtered speech also resulted in greater STG and motor cortex activity.

TABLE 2: COMPLETE LIST OF FOCI USED IN META-ANALYSIS

Study ID	Contrast	Talairach co-ordinates			MNI co-ordinates			Hemisphere	Probabilistic anatomical correlates	Cytoarchitectonic
Parkinson 2012	FAF>noFAF	X	Y	Z	X	Y	Z			
		63	-17	9	56.56	-17.55	11.5	R	Rolandic Operculum	OP1
		66	-19	13	59.26	-19.54	14.76	R	Rolandic Operculum	OP1
		66	-29	9	59.28	-28.44	11.09	R	Superior temporal gyrus	
		51	-28	9	45.54	-27.47	10.91	R	Superior temporal gyrus	TE1.1
		-57	-29	9	-53.4	-27.96	9.34	L	Superior temporal gyrus	TE1.1
		-52	-28	6	-48.79	-26.94	6.97	L	Superior temporal gyrus	
Behroozmand 2015	FAF>noFAF	-51.06	-18.45	3.47	-50.55	-17.71	4.04	L	Superior temporal gyrus	TE1.0
	(ROI analysis)	48.85	-24.73	6.36	48.36	-23.65	6.98	R	Superior temporal gyrus	TE1.1
		45.12	-26.75	7.91	44.67	-25.53	8.5	R	Superior temporal gyrus	TE1.1
Tourville 2008	FAF>noFAF	-22.31	-46.89	62.52	-22	-42	70	L	Postcentral gyrus	SPL (5L)
	(Fixed effects)	42.86	-15.15	43.2	48	-10	44	R	Precentral gyrus	Area 4a
		54.14	8.03	36.58	60	14	34	R	Precentral gyrus	
		63.52	-5.71	22.83	70	-2	20	R		
		52.36	8.56	31.2	58	14	28	R	Inferior frontal gyrus (pars Opercularis)	BA44
		50.6	25.69	29.19	56	32	24	R	Inferior frontal gyrus (pars Triangularis)	
		-51.22	-25.09	13.65	-54	-24	14	L	Supramarginal gyrus	OP1
		-62.45	-38.77	19.37	-66	-38	22			
		-56.73	-30.3	9.46	-60	-30	10	L	Superior temporal gyrus	
		-58.57	-24.7	9.95	-62	-24	10	L	Superior temporal gyrus	TE3
		-56.81	-60.11	6.63	-60	-62	10	L	Middle temporal gyrus	
		61.61	-37.2	18.01	68	-36	18	R	Superior temporal gyrus	IPC (PF)
		65.37	-40.42	12.36	72	-40	12	R		
		61.77	-17.7	10.85	68	-16	8	R	Superior temporal gyrus	
		43.47	-8.75	-3.02	48	-8	-8	R	Superior temporal gyrus	TE3

		50.82	-11	0.49	56	-10	-4	R	Superior temporal gyrus	
		23.53	-54.95	-49.18	26	-62	-54	R	Cerebellum (Lobule VIIIa)	
		9.41	-83.56	32.55	9.32	-79.38	33.75	R		
Zheng 2013	FAF & MASK > all	-32.31	-40.5	-21.73	-31.99	-40.29	-16.53	L	Fusiform gyrus	
	(MVPA)	37.32	-55.24	30.3	36.95	-52.05	30.38	R	Inferior parietal lobule	hIP1
		7.73	9.08	46.71	7.65	11.06	42.5	R	Middle cingulate cortex	
		26.96	-53.33	-27.34	26.69	-52.99	-20.7	R	Cerebellum (Lobule VI)	
Zarate 2010	FAF>noFAF	58.09	-28.51	6.16	64	-28	4	R	Superior temporal gyrus	
		59.88	-23.45	12.08	66	-22	10	R	Superior temporal gyrus	TE3
		0.41	13.37	41.58	2	20	40	R	Middle cingulate cortex	
		0.35	11.16	44.98	2	18	44	L	Supplementary motor area	BA6
		39.15	-5.98	45.81	44	0	46	R	Precentral gyrus	
		28.62	18.4	8.31	32	22	2			
		46.84	5.21	27.18	52	10	24	R	Inferior frontal gyrus (pars Opercularis)	
		35.32	-47.13	43.65	40	-44	48	R	Inferior parietal lobule	IPC (PF)
		52.02	-39.59	42.85	58	-36	46	R	Inferior parietal lobule	hIP2
		-5.35	-2.72	52.57	-4	4	54	L	Supplementary motor area	BA6
		-45.94	-5.19	40.84	-48	0	42	L	Precentral gyrus	
		-32.46	18.72	7.3	-34	22	2	L		
		-49.48	3.32	30.78	-52	8	30	L	Precentral gyrus	
		-40.53	-42.84	40.97	-42	-40	46	L	Inferior parietal lobule	BA2
		-35.02	-50.49	42.14	-36	-48	48	L	Inferior parietal lobule	hIP1
Zheng 2010	Interaction	-62.38	-44.68	12.5	-61.76	-42.68	13.54	L	Superior temporal gyrus	
	between									

	MASK>noMASK and listen>NoMASK	48.98	-18.27	-2.03	48.49	-17.8	-0.94	R	Superior temporal gyrus	TE1.1
Hashimoto 2003	DAF>noDAF	-56.86	-31.26	19.27	-60	-30	21	L	Superior temporal gyrus	IPC (PFop)
		-54.24	-35.11	29.76	-57	-33	33	L	Supramarginal gyrus	IPC (PF)
		-59.65	-20.32	22.96	-63	-18	24	L	Postcentral gyrus	IPC (PFop)
		54.42	-19.08	6.09	60	-18	3	R	Superior temporal gyrus	
		54.11	-32.35	26.45	60	-30	27	R	Supramarginal gyrus	IPC (PFcm)
		59.79	-23.21	19.3	66	-21	18	R	Supramarginal gyrus	OP1
Takaso 2010	Parametric response to delay increase	-58.56	-28.25	7.82	-62	-28	8	L	Superior temporal gyrus	
		-51.04	-9.13	4.35	-54	-8	2	L	Superior temporal gyrus	TE1.2
		-49.38	-39.83	10.48	-52	-40	12	L	Superior temporal gyrus	
		61.7	-36.5	10.87	68	-36	10	R	Superior temporal gyrus	IPC(PF)
		48.86	-21.01	6.72	54	-20	4	R	Superior temporal gyrus	TE1.0
Fu 2006	FAF>noFAF	-32	30	-13	-31.68	28.43	-12.17	L	Inferior frontal gyrus (pars Orbitalis)	
		-28	30	-7	-27.72	28.73	-7.13	L	Inferior frontal gyrus (pars Orbitalis)	
		-53	-26	-7	-52.47	-25.53	-4.78	L	Middle temporal gyrus	
		-21	-73	-7	-20.79	-71.06	-2.81	L	Lingual gyrus	hOC4v (V4)
		-11	23	-2	-10.89	22.19	-2.64	L	Caudate nucleus	
		32	23	-2	31.68	22.19	-2.64	R		
		-61	-23	-2	-60.39	-22.38	-0.71	L	Middle temporal gyrus	
		4	-7	-2	3.96	-6.88	-1.38	R	Thalamus- temporal	
		-4	-76	-2	-3.96	-73.73	1.66	L	Lingual gyrus	BA17
		-57	-4	4	-56.43	-3.68	3.86	L	Superior temporal gyrus	
		36	-67	4	35.64	-64.72	6.76			
		-57	-17	4	-56.43	-16.28	4.46	L	Superior temporal gyrus	TE1.2

		-11	-56	4	-10.89	-54.06	6.25	L	Calcarine gyrus	
		-21	-4	4	-20.79	-3.68	3.86	L		
		-50	-56	4	-49.5	-54.06	6.25	L	Middle temporal gyrus	
		-53	-4	9	-52.47	-3.44	8.45	L	Rolandic Operculum	OP4
		11	-67	9	10.89	-64.48	11.35	R	Calcarine gyrus	BA17
		11	-67	15	10.89	-64.18	16.87	R	Calcarine gyrus	BA18
		-32	-13	15	-31.68	-11.87	14.38	L	OP3	
		-57	-17	20	-56.43	-15.5	19.16	L	Postcentral gyrus	OP1
		-7	-52	26	-6.93	-49.12	26.28	L	Posterior cingulate cortex	
Christoffels 2007	MASK>noMASK	51	-18	8	50.49	-17.05	8.18	R	Heschl's gyrus	TE 1.0
		57	-30	12	56.43	-28.48	12.41	R	Superior temporal gyrus	
		-41	-28	7	-40.59	-26.79	7.72	L	Superior temporal gyrus	TE1.1
Niziolek 2013	FAF>noFAF	-38.5	15.3	7	-38.12	15.16	5.73	L		
		8.8	28.9	41.2	8.71	30	36.53	R	Middle cingulate cortex	
		50.2	27.5	8	49.7	27.03	6.09	R	Inferior frontal gyrus (pars Triangularis)	BA45
		31.4	24.7	9	31.09	24.37	7.13	R		
		-48.1	-16.5	0.3	-47.62	-15.97	1.03	L	Superior temporal gyrus	TE1.0
		-59.8	-44.1	13.8	-59.2	-42.06	14.71	L	Superior temporal gyrus	
		52.8	-5.6	-5.5	52.27	-5.69	-4.38	R	Superior temporal gyrus	TE1.0
		51.8	-20.5	3.9	51.28	-19.67	4.53	R	Superior temporal gyrus	TE1.0
McGuire 1997	FAF>noFAF	-50	-10	0	-49.5	-9.69	0.46	L	Superior temporal gyrus	TE1.0
		46	-20	4	45.54	-19.18	4.6	R	Heschl's gyrus	TE1.0
		-46	-6	8	-45.54	-5.43	7.63	L	Rolandic Operculum	
		-52	-36	16	-51.48	-34.1	16.36	L	Superior temporal gyrus	IPC(PFcm)
Christoffels 2011		52	-20	10	51.48	-18.89	10.11	R	Heschl's gyrus	TE1.0
		43	-28	13	42.57	-26.5	13.23	R	Heschl's gyrus	OP1

	Parametric response to noise level	-47	-30	9	-46.53	-28.63	9.65	L	Superior temporal gyrus	TE1.1
Toyomura 2007	FAF>noFAF	61.38	-20.25	21.23	62	-22	22	R	Supramarginal gyrus	OP1
		55.44	15.12	28.76	56	14	32	R	Inferior frontal gyrus (pars Opercularis)	BA44
		35.64	28.97	-2.94	36	30	-2	R		
		-51.48	9.59	34.55	-52	8	38	L	Precentral gyrus	BA44
		51.48	-9.79	-1.26	52	-10	-2	R	Superior temporal gyrus	TE1.0
		31.68	-38.56	42.36	32	-42	44	R	BA2	

3.7. ACTIVATION LIKELIHOOD ESTIMATION ANALYSIS

To assess the convergence in feedback regions across studies, an activation likelihood estimation (ALE) analysis was carried out. Peak co-ordinates resulting from the altered>unaltered feedback contrast in each study were selected for inclusion, unless a study used an analysis that explicitly compared altered feedback with both unaltered feedback and listening (for example, Zheng et al's 2009 interaction analysis). In that case, those co-ordinates were used instead; the aim was to select co-ordinates that were the best available evidence of a feedback response in each study. One study, Hirano et al. (1997) was excluded from the analysis as the altered feedback was compared with rest rather than a speech condition.

Co-ordinates given in Talairach space were converted to MNI space for the meta-analysis, using Brett et al.'s (2001) transform. An ALE meta-analysis was carried out using a non-additive random effects model, as described by Eickhoff et al. (2009), revised by Turkeltaub et al. (2012) and implemented in GingerALE software version 2.3.6 (www.brainmap.org). For each voxel, activation likelihood estimates were calculated by modelling each co-ordinate using a 3-D Gaussian probability density function with a FWHM determined by the number of subjects in each study (median FWHM 9.8 mm, range 9.2-10.9). Study-specific activation probabilities were merged to create an ALE statistic at each voxel; the resulting ALE map was then corrected for multiple statistical comparisons using a voxelwise threshold of FWE $p<0.05$, as recommended by Eickhoff (2009). There was no minimum cluster size. The results were projected onto a standard template in MNI space.

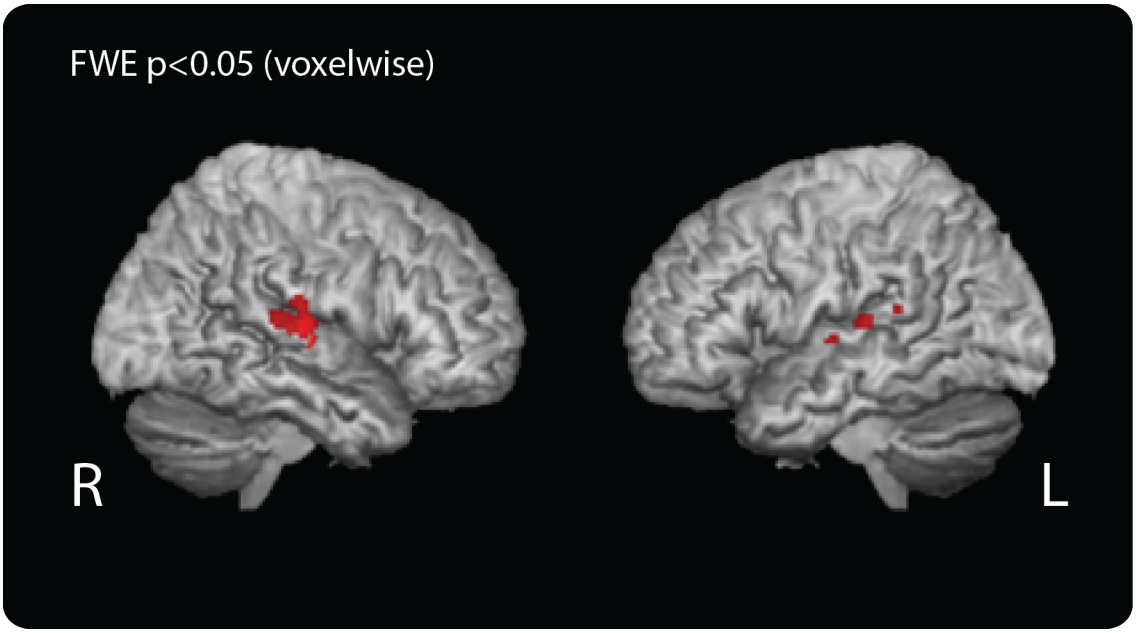


FIGURE 8: REGIONS OF SIGNIFICANT CONVERGENCE BETWEEN ACTIVATION FOCI IN THE 14 SELECTED AUDITORY FEEDBACK PERTURBATION STUDIES, AS REVEALED BY AN ACTIVATION LIKELIHOOD ESTIMATION ANALYSIS. CORRECTED FOR MULTIPLE COMPARISONS USING FWE. P<0.05

The ALE analysis revealed four significant clusters, one in the right hemisphere and three in the left (Figure 8). In both hemispheres, activation spanned superior and transverse temporal gyri, while in the right hemisphere activation also encompassed precentral and postcentral gyri, and insula. 18 foci from twelve studies contributed to the first cluster, in the right hemisphere. Fewer studies overlapped in the left hemisphere, with a total of six foci from six different studies contributing to three significant clusters. Cluster extent and co-ordinates of ALE extrema are given in Table 3, along with probabilistic anatomical labels for each extrema derived automatically by the GingerALE software. In the right hemisphere, there were three extrema, in STG, transverse temporal gyrus, and pre central gyrus. In the left hemisphere, three extrema were found in STG, and one in the transverse temporal gyrus. With the exception of one peak in the left hemisphere, activation in the STG was seen in posterior, rather than anterior regions.

TABLE 3: RESULTS OF ALE META-ANALYSIS

Cluster #	Foci (studies)	Volume (mm^3)	Weighted centre (x,y,z)			Extrema Value	Extrema (x, y, z)			Hemisphere	Macroanatomical label	Cytoarchitectonic label
1	18(12)	2704	52.8	-20.2	8.4	0.03511154	52	-18	8	Right	Transverse temporal gyrus	BA 41
						0.02104977	60	-20	14	Right	Postcentral gyrus	BA 40
						0.019313881	58	-28	10	Right	Superior temporal gyrus	BA 41
2	3(3)	336	-52.9	-27.6	9.7	0.019327119	-54	-28	10	Left	Superior temporal gyrus	BA 41
						0.016157601	-48	-28	10	Left	Transverse temporal gyrus	BA 41
3	1(1)	88	-50.4	-14.7	2.7	0.016413487	-50	-14	2	Left	Superior temporal gyrus	BA 22
4	2(2)	40	-60.4	-41.2	14.8	0.016533166	-60	-42	14	Left	Superior temporal gyrus	BA 22

3.8. SUMMARY OF EVIDENCE

IS THE EVIDENCE FOR THE ROLE OF STG IN FEEDBACK MONITORING ROBUST AND CONCLUSIVE?

There are multiple components to successfully demonstrating perceptual modulation in auditory cortex during speech production. First, any neural activation should be associated with behavioural evidence that the subject has adapted to the feedback manipulation. Second, neural responses in the putative feedback control area should be attenuated when subjects speak with unmanipulated feedback compared to hearing stimuli without articulating. Finally, the feedback control region should be more active when the subject speaks with altered feedback than when they speak normally. All but one of the studies included in this review fulfilled the last criterion, but many did not tackle the first or second question, and those that did found variable results. Although those studies that reported behavioural adaptation results for FAF generally found some evidence of compensation for the frequency shift, this was often a very small proportion of the total frequency shift- in the most extreme example altering their voice on average by 4.5 cents to compensate for a shift of 600 cents—meaning that there would still be a mismatch between auditory feedback and targets. There are some characteristics of the way we perceive our voices through bone conduction that mean that talkers can have difficulty accurately matching the loudness or pitch of an external stimulus (Murry, 1990) so we should not expect perfect compensation for manipulations. Nevertheless the behavioural adaptation found in these studies is quite far short of what talkers are capable of. For example, Zarate et al (2010) found that when asked to deliberately compensate for a frequency shift, talkers were capable of compensating for 87.6% of a 200 cent frequency shift and over-compensated for a 25 cent shift. Although their participants were trained singers and were therefore likely to perform better at this task

than non-singers, non-singers are also able to make considerable adjustments when prompted to do so (Murry, 1990).

Delayed auditory feedback is more difficult to measure adaptation to, since there is no way to compensate for a delay in the same way that a pitch shift can be opposed. Instead, studies measured disruption to speech, using either objective measures (calculations of speech rate) or subjective measures (self- ratings of speech difficulty and quality). According to these measures, DAF caused subjects' speech to become less fluent and more effortful (Hashimoto & Sakai, 2003).

Two studies (Christoffels et al., 2007; 2011) explicitly instructed participants to avoid any kind of adaptation to the perturbation (masking noise). This was intended to keep the signal-to-noise ratio constant and thus ensure that the masker was equally effective across all trials. Although Christoffels et al. (2007; 2011) reported that participants were able to keep their vocal intensity constant across all speech conditions, because this involves suppressing the natural behavioural response to noise, the cognitive effort involved in this suppression may confound the neural results.

The second criterion for identifying a feedback response is the presence of speaking-induced suppression during normal speech production. That is, responses in the supposed feedback region should be lower during overt speech with normal feedback than when listening to a comparable sound. Over half the studies discussed here did not include a listening condition, so were unable to confirm the presence of speaking-induced suppression. Of those that did include a listening condition, three (Behroozmand et al., 2015; McGuire et al., 1996; Christoffels et al, 2007) found no significant differences in the STG BOLD responses to listening and to speaking with unmanipulated feedback. One, Zarate et al. (2010) found that STG responses were actually higher when vocalizing alone

than when listening. In total, then, only three studies (Christoffels et al., 2011; Zheng et al., 2009; Zheng et al., 2013) found evidence of speaking-induced suppression, while twelve either found no evidence or did not look for it.

The final criterion for defining a feedback control region is that it should respond more to feedback 'error' or perturbation than to unperturbed feedback. All but one study (Hirano et al., 1997) made this comparison. They found activation in middle and superior temporal cortex, inferior frontal gyrus and pre central gyrus amongst other regions. An ALE meta-analysis of foci from all 14 studies that included a perturbation vs no perturbation contrast showed that activation foci tended to overlap in STG and in precentral gyrus. However, this included correlates from five studies that did not correct for multiple comparisons, so results should be interpreted with caution. The strength of the ALE is also limited by the relatively small number of studies in this area, although a more stringent FWE correction was applied to compensate for this.

WERE THE TASKS APPROPRIATE TO THE INVESTIGATION?

Since it is difficult to reliably elicit errors in typical fluent speech, or to measure compensation for error, it has been necessary for studies to introduce external perturbations. Nevertheless, since the aim of the research is to draw conclusions about the mechanisms behind speech as humans typically use it, it is desirable that the perturbations have some kind of relation to situations that talkers might encounter in everyday life. Masking noise might be considered the most ecologically valid approach, since most people will experience a conversation in a noisy environment outside the lab, whereas they are unlikely ever to hear the pitch of their voice shift suddenly unless they make a habit of inhaling helium, and delayed auditory feedback is rarely heard outside of faulty phone lines or recording booths. That said, the subtle pitch shifts that affect only

the first and second formants mimic the kind of misarticulating that the error correction mechanism is supposed to deal with, with both the HSF (Hickok, 2014b) and DIVA (Guenther, 2006) models stating that auditory targets are likely defined at the phoneme level. Exactly what the 'error' is that participants experience during other forms of feedback perturbation are ill-defined. In the case of frequency-altered feedback, the target may be utterance-level or phoneme-level pitch, depending on the type of manipulation. Presumably the target when participants speak in noise is utterance-level intensity, pitch and spectral centre of gravity, since these are the acoustic features that participants most often change in compensation (Cooke & Lu, 2010). For delayed auditory feedback, it may be co-ordination of somatosensory with auditory feedback. This seems to be getting farther and farther away from Hickok's and Guenther's definition of an auditory target, and yet all three types of study have been cited as evidence in support of each model's conceptualisation of the feedback loop.

A secondary problem is that many of the experimental procedures used require very tightly constrained speech tasks, which may not fully represent the ways in which talkers typically use speech. The most extreme example is that four studies required participants to articulate the phoneme /a/ for up to five seconds, dozens of times per experiment. This is closer to singing or to gargling for the doctor than it is to anything in normal fluent speech. Other studies used whole words, but again in many cases the stimuli set was highly restricted, with as few as eight different monosyllabic word stimuli, and one study (Zheng et al., 2013) apparently using just a single stimulus word ('Ted'), repeated 72 times per functional run. The use of such a small number of stimuli in this and in other studies may have led to semantic satiation (Smith & Klein, 1990) Although it appears that semantic content is not necessary to prompt adaption, since other studies used

meaningless phonemes, the possibility of a satiation effect does add a potential confound to the experiment design, and may have caused neural responses to be attenuated (Pilgrim, Fadili, Fletcher, & Tyler, 2002). Only three studies used connected speech (sentences) as stimuli. In speech perception, unconnected speech such as single words are processed differently to whole sentences (Peelle, 2012), so it is possible that in speech production, the activation we see is dependent on the type of task used.

Fu et al.'s (2006) finding that talkers were more likely to misattribute their speech to an external source when they heard it pitch-shifted is intriguing because it suggests an alternative explanation for the response in STG- if altered feedback were perceived as a sound that was not self-produced, this might explain a release from speaking-induced suppression. Nevertheless, the fact that participants compensated for shifts suggest that they did treat the altered feedback as their own voice.

CONCLUSIONS

All studies found activation of some kind in STG, and an activation likelihood estimation analysis found significant overlap between experimental foci in STG, transverse temporal gyrus and precentral gyrus. However, evidence that the STG functions as an error monitor remains inconclusive, as most experiments were not designed to rule out confounds, such as the possibility that activation was related to hearing unusual sounds in the manipulated feedback condition, and not registering this as 'error'. Moreover, neural responses varied in strength across studies, with many experiments failing to find results at the whole brain level when corrected for multiple comparisons. Some persuasive evidence for the STG as an error monitor comes from studies that found that behavioural adaptation correlated with activation in STG. However, none of the studies considered here demonstrated all three components of a feedback response- that is, behavioural

adaptation, speaking-induced suppression, and an STG response to altered feedback. It is important that future research includes both a listening control condition and some measure of behavioural compensation for perturbation, so that confident conclusions about the role of STG in error monitoring and speech production may be drawn.

CHAPTER 4: A CASE STUDY OF SPEECH FEEDBACK CONTROL AFTER STROKE

4.1. ABSTRACT

The previous chapter has looked at responses to auditory feedback perturbations in neurotypical speakers. Here, we ask if lesions to the STG affect the ability to respond appropriately to altered feedback. This chapter is a report describing the case of a 46-year-old man who reported being unable to hear his own voice clearly following a left middle cerebral artery infarct involving temporal cortex. Testing showed that he had no hearing impairment, and his perception of other sounds was unaltered. This study used a speech production in noise task to investigate whether the patient's perceptual issues had affected his ability to use auditory feedback. The subject spoke in three different levels of masking noise. Comparison to neurotypical controls revealed that the patient over-compensated for the noise, raising vocal intensity and pitch more than controls across all three levels of masking- although, because of variability in the control group, and a small sample size, this did not reach statistical significance. However, there was a significant difference between the patient and controls in the percentage of unvoiced frames, indicating that the patient had more difficulty speaking in noise than controls. Results were consistent with attenuated perception of his own voice compared to external sounds. Other stroke patients show an impaired ability to repair errors in their own speech, meaning that this may be a manifestation of a more general problem.

4.2. INTRODUCTION

This chapter discusses an unusual self-monitoring impairment reported by a patient experiencing expressive aphasia as a result of stroke. This section provides an overview of the causes and consequences of stroke, discusses techniques that have previously been used to assess the use of auditory feedback in patients with aphasia, and offers arguments for using speech production in noise as an assessment tool in preference to other feedback manipulation techniques.

CAUSES AND CONSEQUENCES OF STROKE

A stroke occurs when a lack of blood flow to the brain results in cell death. There are two types of stroke- ischaemic stroke, in which a thrombus (blood clot) blocks a blood vessel (known as an infarction), and haemorrhagic stroke, in which blood vessel dissection leads to blood leaking and damaging surrounding brain tissue; both types of stroke are most commonly caused by hypertension, or high blood pressure. The effect of the stroke is largely determined by the vascular territory in which the infarction or haemorrhage occurs.

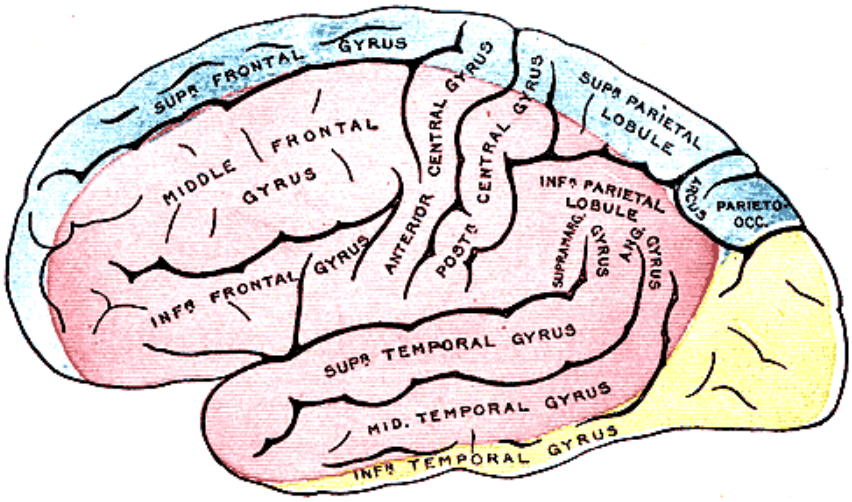


FIGURE 9: OUTER SURFACE OF CEREBRAL HEMISPHERE, SHOWING VASCULAR TERRITORIES (GRAY, 1918). BLUE AREAS ARE SUPPLIED BY THE ANTERIOR CEREBRAL ARTERY, YELLOW BY THE POSTERIOR CEREBRAL ARTERY, AND RED BY THE MIDDLE CEREBRAL ARTERY.

The basilar artery supplies blood to the posterior portion of the brain, including the brain stem and cerebellum, while the carotid arteries supply the anterior portion of the brain. Blood is supplied to the cerebrum by the anterior, posterior and middle cerebral arteries. The posterior cerebral artery branches from the top of the basilar artery and covers the occipital lobe as well as inferior and medial temporal lobes. The anterior and middle cerebral arteries (ACA and MCA) both branch from the internal carotid artery; the ACA supplies medial frontal and parietal cortex, while the MCA covers the rest of frontal, parietal and temporal cortex as well as the basal ganglia and internal capsule. The effects of an ischaemic stroke are typically limited to the vascular territory in which it occurs, while haemorrhagic strokes can cross territory boundaries.

As the MCA is the largest vessel branching off the internal carotid artery, and covers the greatest portion of the brain, most strokes occur in MCA territory. Occlusion of the MCA stem can lead to severe disability or death, while strokes affecting its branches cause a wide variety of deficits, as the MCA supplies so many different parts of the brain. Critically, both the posterior superior temporal gyrus and inferior frontal gyrus are part of MCA territory, meaning that MCA strokes affecting the dominant hemisphere for language can often cause aphasia.

TYPES OF APHASIA

Two common types of aphasia are often caused by strokes that affect the middle cerebral artery: expressive, or 'Broca's' aphasia, which is associated with occlusions affecting the upper division of the left MCA, and receptive ('Wernicke's') aphasia, which typically results from lesions involving regions supplied by the inferior division of left MCA. 34-38% of stroke patients experience aphasia after stroke (Bakheit, Shaw, Carrington, & Griffiths, 2007), of which 12% are likely to be expressive aphasics and 16% receptive

aphasics; global aphasia, which combines both types of deficit, accounts for 32% of aphasic patients (Pedersen, Vinter, & Olsen, 2004). Broca's aphasia gets its name from the eponymous Paul Broca, who is credited with discovering a link between neurological damage to the left hemisphere and aphasia- although the link had actually been made some 25 years earlier by Marc Dax (Buckingham, 2006). Wernicke's aphasia similarly derives its name from Carl Wernicke, who linked receptive aphasia with lesions in the posterior superior temporal gyrus. Both Broca's and Wernicke's aphasia are terms that have been applied to disorders resulting from a wide range of lesions, which do not always affect Broca's or Wernicke's areas as classically defined. Indeed, there is some disagreement about how much of the posterior STG Wernicke's area actually includes (Bogen & Bogen, 1976). The two types of aphasia are here referred to by the descriptive terms 'expressive' and 'receptive' aphasia, to avoid assumptions about the lesions underlying the deficit.

Receptive aphasia is characterised by syntactically well-formed speech that may contain neologisms and make little sense semantically. People with receptive aphasia are typically unable to understand written or spoken language, including their own speech. By contrast, expressive aphasia is characterised by dysfluent speech production contrasted with relatively unimpaired comprehension. Non-fluency in this context includes problems with articulation and prosody as well as agrammatic speech (though not all receptive aphasics have deficits in each of these areas). Additionally, patients with expressive aphasia may have difficulty understanding some syntactically complex sentences, although words and simple sentences are usually readily understood. Broca identified a region in the inferior frontal gyrus, more specifically the inferior frontal operculum (Brodmann areas 44 and 45), which is thought to be associated with

expressive aphasia and is known as Broca's area. However, damage that only affects Broca's area can result in apraxia of speech (a deficit in motor planning and execution of speech) without any other symptoms of 'Broca's' aphasia (Trupe et al., 2013). It has therefore been suggested that 'Broca's' aphasia is a vascular syndrome, resulting from loss of function in a wider network of areas supplied by the upper division of the left MCA, including Broca's area, the insula, and surrounding cortex (Mohr et al., 1978). Additionally, research has found that while an infarct affecting the pars opercularis is not in itself sufficient to cause Broca's aphasia, patients with lesions in both pars opercularis and the left superior temporal gyrus were highly likely to exhibit Broca's aphasia, to the extent that it was possible to predict expressive aphasia with 95% accuracy based on proportional damage to these two areas (Fridriksson, Fillmore, Guo, & Rorden, 2015). This suggests a link, rather than a dissociation, between the two areas and their role in speech production.

SELF-MONITORING AFTER STROKE

Do stroke patients monitor and correct errors in their speech differently to controls? Studies looking at self-monitoring abilities after stroke have focused on three areas: covert repairs (in which the patient begins to make a mistake and then corrects it), overt repairs (in which the patient corrects something they have just said) and the ability to cope with adverse feedback conditions (i.e. delayed auditory feedback). Overt repairs and responses to delayed auditory feedback are assumed to rely on postarticulatory monitoring processes, or feedback control, while covert repairs or 'prepairs' (Schlenck, Huber, & Willmes, 1987) are presumed to reflect the use of prearticulatory monitoring, or feedforward control. Patients with both expressive and receptive aphasia make more semantic and phonological errors than controls, but are less likely to correct themselves

(Oomen, Postma, & Kolk, 2001; Schlenk et al., 1987). This is perhaps to be expected in patients with receptive aphasia, who are characteristically unaware that their speech contains errors (Weinstein, Lyerly, Cole, & Ozer, 1966), but unexpected in expressive aphasics, who are usually considered to be aware that their speech is agrammatic and effortful (Kertesz & McCabe, 1977). Research confirms that even aphasic patients with preserved auditory comprehension tend not to correct or demonstrate awareness of speech errors (Maher, Rothi, & Heilman, 1994, Oomen et al., 2001), although Schlenk et al. (1987) found that patients with high auditory comprehension skills made more covert, but not overt, repairs. Intriguingly, this apparent monitoring deficit seems only to apply to self-produced speech, since patients with expressive aphasia are able to identify and correct a high proportion of speech errors when asked to identify semantic and phonological slips in the speech of others (Oomen et al., 2001).

Oomen et al. (2001) suggest that patients with expressive aphasia rely more on prearticulatory monitoring than postarticulatory monitoring, based on an experiment where patients spoke in quiet and in 90dB of white noise. The noise was presumed to mask auditory feedback of the participant's voice and make postarticulatory monitoring unavailable. While controls corrected more errors when they spoke in quiet (when auditory feedback was available) than in noise (when it was not), there was no significant change in the number of errors corrected between the two conditions in patients with aphasia. This is to be expected if patients rely more on prearticulatory monitoring processes, so the presence of auditory feedback makes little difference to how many errors they correct. This theory is apparently also supported by the finding that aphasic patients make more covert repairs (which are presumed to rely on feedforward

monitoring) than overt repairs, which are assumed to result from feedback monitoring (Oomen et al., 2001, Schlenk et al., 1987).

However, this conclusion is apparently contradicted by a group of studies looking at the effects of delayed auditory feedback on patients with aphasia— which seem to indicate that aphasic patients are more affected by auditory feedback perturbations than controls. Under delayed auditory feedback (which is presumed to disrupt speech by preventing effective feedback monitoring (Yates, 1964), patients with expressive aphasia made more phonemic errors and spoke with a longer duration than controls with dysarthria or a learning disability (Singh & Schlanger, 1969). Additionally, in a study comparing both non-fluent and fluent aphasics with neurotypical controls, one study (Boller, Vrtunski, Kim, & Mack, 1978) found that non-fluent aphasics were significantly more affected by DAF than controls, while fluent aphasics were less affected by DAF than controls or non-fluent aphasics. Another found that patients with a left hemisphere lesion were more disrupted by DAF when completing simple verbal tasks (such as counting from 1 to 10) than controls, but there was no difference between left-hemisphere stroke patients and controls when the task was a non-verbal one such as finger-tapping. Conversely, patients with a right-hemisphere lesion were more disrupted by DAF than controls during non-verbal rhythmic tasks, but did not significantly differ from controls in the verbal task (Vrtunski, Mack, Boller, & Kim, 1976). This suggests that patients with left-hemisphere stroke have a specific difficulty with speech-related feedback monitoring rather than rhythmic movement- once again conflicting directly with the assumption that patients with expressive aphasia tend to ignore speech feedback and rely primarily on feed-forward monitoring. Evidence for the importance of feedback versus feed-forward monitoring in patients with expressive aphasia thus remains mixed.

SPEECH IN NOISE AS DIAGNOSTIC TOOL

One difficulty with interpreting the research so far is that the effects of delayed auditory feedback are hard to quantify, as there is considerable individual variability in responses to DAF even in typical speakers: it may cause speech rate to increase, slow down or stop entirely (Yates, 1963). To address this, here we used the adaptation to masking noise, known as the Lombard response (Lombard, 1911), as an index of the extent to which this patient could respond appropriately to auditory feedback manipulation. In contrast to DAF and frequency altered feedback, both techniques that do not always elicit vocal adaptation (Burke, 1975; Lametti et al., 2012), masking noise reliably elicits adaptation in all speakers, to the extent that it was initially proposed as a way to identify malingerers who were pretending to be deaf (Lombard, 1911).

There are several reliable acoustic correlates of Lombard speech, which make the degree of adaptation easily quantifiable. In this study, adaptation was measured using three of the most commonly identified acoustic characteristics of the Lombard response: vocal intensity, pitch and distribution of spectral energy. A brief recapitulation of the research follows, to outline what vocal changes are expected in healthy talkers.

The most immediately obvious correlate of Lombard speech is an increase in vocal intensity or amplitude. The magnitude of the increase is partly dependent on the background noise level, and increases in direct proportion to the level of the masker (Dreher & O'Neill, 1957; Webster & Klumpp, 1962). However, there is also considerable variation between speakers and tasks. Studies measuring increases in vocal intensity in masking noise relative to quiet have found changes ranging from an increase of 5.6 dB SPL in response to 90 dB SPL of white noise (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988), to 23 dB SPL in 85 dB of masking noise (Webster and Klumpp, 1962).

Similar increases have been found for pitch, which is unsurprising as increasing sub-glottal pressure can increase the rate of vibration of the vocal folds as well as causing a concomitant increase in intensity (Lu & Cooke, 2008; Plant & Younger, 2000). However, it is sometimes difficult to compare between experiments, as there seems to be little agreement on the most appropriate measure of pitch to use; studies have analysed mean F0 in semitones (Lu & Cooke, 2008), peak F0 in Hz (Patel et al., 2008) and mean vowel F0 (Junqua, 1993), amongst a multiplicity of other methods. Nonetheless, despite heterogeneity of analysis parameters, most studies conclude that an increase in masker intensity is associated with a corresponding increase in pitch. Finally, Lombard speech is characterised by changes in the distribution of energy across the spectrum. There are a number of ways that this can be measured. In this study, energy distribution was measured using the spectral centre of gravity (CoG). This is the frequency which divides the spectrum into two, such that the amount of energy in both parts is equal. Hence, a sound that has a low spectral CoG has more energy in the lower frequencies than the high frequencies, whereas the reverse is true for sounds with a high CoG. However, there are other similar measures of energy distribution; for example, the slope of the speech spectrum, or spectral tilt. A flattening of spectral tilt, in the context of Lombard speech, means that the distribution of energy has changed to increase the contribution of high frequency components. Finally, spectral dispersion, or standard deviation, measures whether the energy is concentrated mainly around the centre of gravity, or spread out over a range of frequencies. Across these different measures, studies agree that Lombard speech is characterized by an energy shift to higher frequencies, both at utterance level (Webster and Klumpp, 1962; Junqua, 1993; Lu & Cooke, 2008; Vadarajan & Hansen, 2006; Tartter et al, 1993) and at phoneme level (Lu & Cooke, 2008).

COMMUNICATIVE STRATEGY OR REFLEX?

It seems reasonable to suggest that the Lombard effect has some role in facilitating communication. Most obviously, raising vocal intensity directly mitigates the effect of noise by increasing the signal to noise ratio (SNR). This adjustment does not completely compensate for changes in the SNR due to noise, however (Lane & Tranel, 1971). Lane, Tranel and Sisson (1970) concluded that this is owing to the disparity between the perceived loudness of one's own voice (the autophonic scale) and the perceived loudness of external sounds (the sone scale). That is, speakers will consider themselves to have doubled their vocal level with only 2/3 the increase in sound pressure that it would take for them to consider an external sound to be twice as loud. This means that the level of intensity that the speaker believes sufficient to compensate for changes in the SNR is lower than the level of intensity actually required to keep the SNR constant for the listener. This is important in the context of this study because it suggests that the Lombard response is driven by the perceived signal to noise ratio. If a subject perceives their voice as quieter in relation to the masking noise, they may attempt to increase the SNR by raising their vocal intensity and would therefore over-compensate compared to controls with typical voice perception.

Additionally, Summers et al. (1988) found that when Lombard speech and speech produced in quiet were equated for amplitude and presented at the same SNR ratio, Lombard speech was the more intelligible. This suggests that the other characteristics of Lombard speech that we have observed so far also play a role in making it more intelligible. However, if we take the enhanced intelligibility of Lombard speech as an indication that it is a conscious strategy to optimise communication in noisy environments, it is difficult to reconcile this idea with the qualities of Lombard speech

that initially led to it being identified as a reflex. That is, it does not occur only in communicative situations, but is an automatic response to speaking in noise. Moreover, it is difficult to voluntarily suppress (Pick, Siegel, Fox, & Kearney, 1989). Evidence from animal studies seems to support the idea that Lombard speech can be a purely reflexive action. In decerebrate cats, which owing to the absence of inhibitory influences of the cerebral cortex generally show no ability to voluntarily control their utterances, Lombard vocalizations were observed (Nonaka, Takahashi, Enomoto, Katada, & Unno, 1997). It seems likely, therefore, that Lombard speech is both an automatic response and a communicative strategy. That is, communicative intent is not necessary for Lombard speech to be produced, but nor is the simple presence of noise sufficient to elicit the fullest range of changes. The difference between Lombard speech with communicative and non-communicative intent can be illustrated by comparing the studies of Dreher and O'Neill (1958) and Webster and Klumpp (1962). Both asked speakers to read aloud words and sentences in comparable levels of broadband noise (between 65 and 100 dB), but whilst Webster and Klumpp found a reliable increase in speech level of 5dB for every 10dB increase in masker intensity, Dreher and O'Neill's speakers barely changed their voices to accommodate for increasing noise level, with a 1dB increase in intensity for every 10dB masker increase. The studies were very similar, with overlapping noise levels (65- 85 dB for Webster & KIumpp; 70-100 dB for Dreher & O'Neill) and similar list-reading tasks. The principal difference in set-up was that Webster and Klumpp's experiment was conducted in pairs, with one person reading the word list and the other repeating words back to them. Each pair was told that if they failed to reach 90 per cent accuracy they would have to repeat the task. Thus, speakers were both highly motivated to maintain communicative accuracy, and had feedback on how well they were performing based on whether their partner correctly repeated the word back to them.

Dreher and O'Neill's speakers, by contrast, had no communicative partner and were given no motivation to speak above the noise. It is probable that the difference between the studies demonstrates the difference between the reflexive and the communicative contributions to the Lombard effect.

A more recent study directly compared Lombard speech in communicative and noncommunicative tasks. Garnier, Henrich & Dubois (2010) compared Lombard speech of speakers playing a game involving river names in pairs and by themselves. They measured vocal intensity, F0, vowel duration and centroid of the speech spectrum in conditions of quiet and 85dB SPL babble. Although these acoustic parameters increased from quiet to noise in both communicative and non-communicative conditions, the differences between speech in quiet and speech in noise were always greater when the speaker was interacting with a communicative partner compared to the non-communicative condition.

Overall, then, the Lombard effect does not require a communicative partner to manifest itself. However, the presence of a communicative intent enhances Lombard speech. In this experiment, a communicative element was introduced by seating the participant opposite the experimenter and asking a series of autobiographical questions. To ensure a clear recording, the experimenter was not able to speak while the participant answered the question, but was able to make encouraging non-verbal gestures to convey that they were listening.

4.3. CASE PRESENTATION

We report the case of a 46-year-old right-handed man experiencing altered perception of his own voice following stroke. The patient presented with left middle cerebral artery (LMCA) infarct secondary to a left internal carotid artery dissection, with thrombus in the LMCA. This unusual combination of ischaemia and haemorhage resulted from previously undiagnosed Ehlers-Danlos syndrome, which causes weakening of the connective tissue and has been known to cause spontaneous internal carotic artery dissection in other patients (Schievink, Limburg, Oorthuys, Fleury, & Pope, 1990). The infarct involved the insula, frontal operculum and middle frontal gyrus with extension into the left parietal lobe and the posterior STG; the lesion is shown in Figure 10 below. Stroke deficits included right-sided weakness, right hemianopia (visual neglect), and limb apraxia. His language abilities were assessed using the Comprehensive Aphasia Test (Porter & Howard, 2004). He displayed difficulties with fluent and syntactically correct speech production, contrasted with relatively intact comprehension abilities.

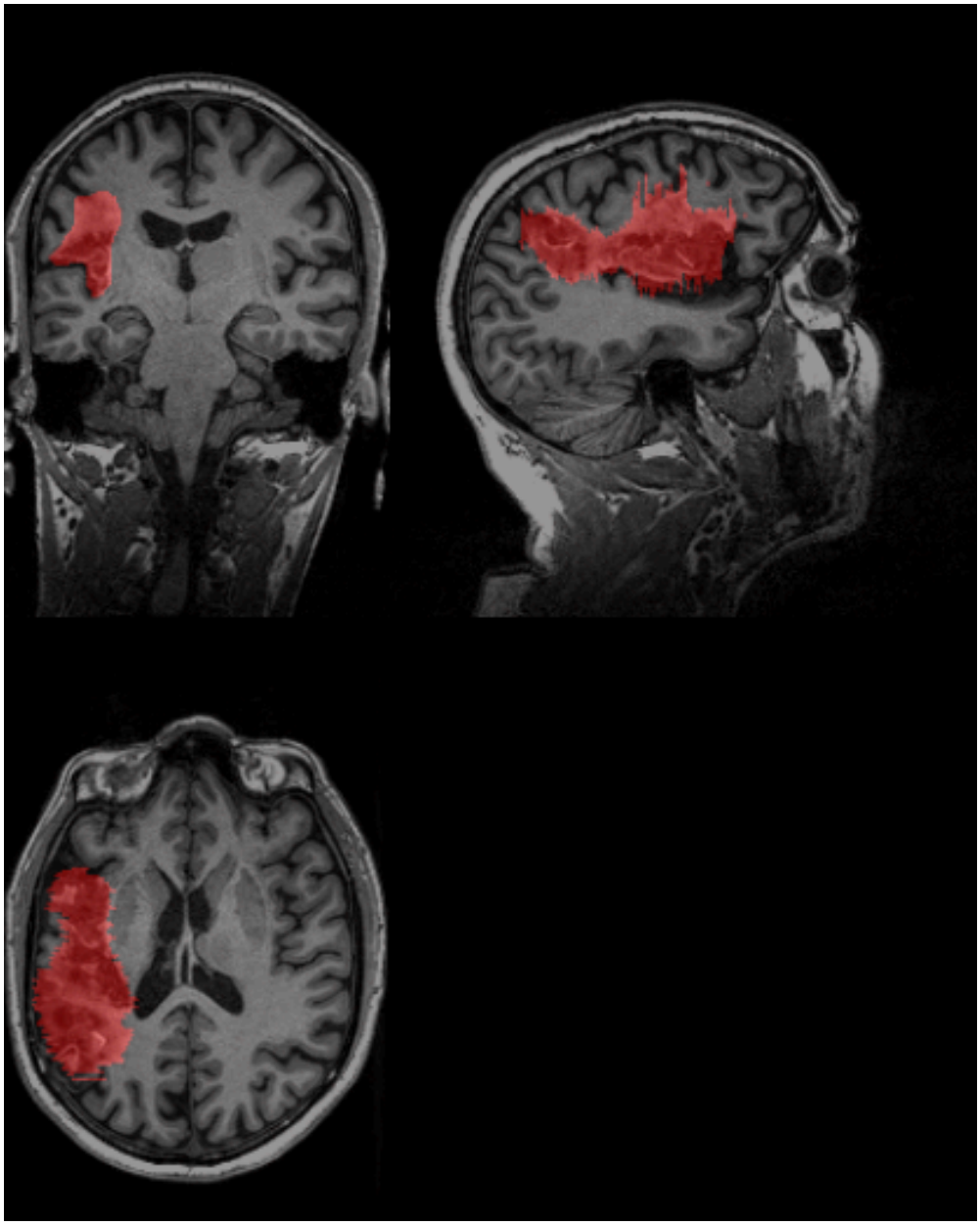


FIGURE 10: T1-WEIGHTED STRUCTURAL SCAN OF PATIENT'S BRAIN, WITH LESIONED AREAS INDICATED IN RED.

The patient reported difficulty hearing his own voice, compared to his perception of external sounds, which remained unaltered. He described his voice as sounding quiet, as if it were under water, far away, or occurring at a delay of up to half a second. The problem began after his stroke, and had persisted for three years at the time he was seen. He reported that some days were better than others, but there appeared to be no consistent aggravating factors. He had previously been a skilled mimic, but reported having lost this

ability since the stroke. He also reported being unable to sing. He had no hearing loss (defined as a four-frequency pure tone average hearing threshold of less than 20dB) and informal assessments showed he had no difficulty perceiving and reacting to external stimuli. He was at ceiling on a task requiring discrimination between sounds.

To assess the impact that his voice perception had on his ability to self-monitor and control his voice, his performance on a speech production in noise task was compared to healthy controls of a similar age with normal hearing.

4.4. PRELIMINARY EVALUATIONS

The patient was referred to our investigative team approximately one and a half years after his stroke. His language abilities had previously been assessed using the Comprehensive Aphasia Test (CAT) shortly after his stroke, and again at four months post stroke. Immediately after his stroke, he was unable to speak, and his spoken and written comprehension were impaired. He underwent two and a half months of rehabilitation treatment (30 minutes, five times a week), as well as 30 hours of speech and language therapy. Four months after his stroke, following the therapy, his comprehension was relatively preserved, while speech production remained below the cut-off score. There was no cognitive impairment. His full CAT scores are shown in the table below.

TABLE 4: PERFORMANCE ON THE COMPREHENSIVE APHASIA TEST

Section	4 days post stroke	4 months post stroke	Cut-off Score
COGNITIVE SCREEN			
Line bisection	0	0	+- 2.5
Semantic memory (/10)	7	10	8/10
Word fluency	0	16	13
Recognition memory (/10)	10	10	8/10
Gesture object use (/12)	9	10	9/12
Arithmetic (/6)	6	5	1/6
Cognitive total (/38)	32	35	
LANGUAGE COMPREHENSION (SPOKEN)			
Comprehension of spoken words (/30)	25	30	25/30
Comprehension of spoken sentences (/32)	10	23	27/32
Comprehension of spoken paragraphs (/4)	4	4	2/4
Spoken comprehension total (/66)	39	57	56/66
LANGUAGE COMPREHENSION (WRITTEN)			
Comprehension of written words (/30)	13	29	27/30
Comprehension of written sentences (/32)	5	25	23/32
Written comprehension total (/62)	18	54	53/66
REPETITION			
Repetition of words (/32)	0	24	29/32
Repetition of complex words (/6)	0	1	5/6
Repetition of nonwords (/10)	0	6	5/10
Repetition of digit strings (/14)	0	6	8/10
Repetition of sentences (/12)	0	6	10/12
Repetition total (/74)	0	43	67/74
NAMING			
Naming objects (/48)	0	38	43/48
Naming actions (/10)	0	6	8/10
Word fluency	0	16	
Naming total	0	60	69

READING ALOUD			
Reading words (/48)	0	24	45/48
Reading complex words (/6)	0	1	4/6
Reading function words (/6)	0	6	3/6
Reading nonwords (/10)	0	5	6/10
Reading total (/70)	0	34	58/70
WRITING			
Copying (/27)	27	27	25/27
Writing picture names (/21)	9	21	15/21
Writing to dictation (/28)	22	26	24/28
Writing total (/76)	58	74	66/76

The patient was interviewed and a range of informal tests were conducted in an attempt to better understand his experience. At the time of testing, he was approximately 18 months post stroke, and his language production and comprehension abilities were consistent with his earlier CAT scores at 4 months post stroke. When given delayed auditory feedback of a live speaker, he described the mismatch between seeing the person speak and hearing their voice as similar to his experience of his own voice. Because he also described his voice as being 'delayed', we considered it possible that he had a more general problem with processing sound- it could be that he perceived all sounds as attenuated or at a delay, but that his own voice was the most obvious. However, he was able to clap in time with a second person, even when he could not see their hands. We probed this further with a computer based test, that required him to hear and make fine temporal distinctions between sounds. In this task, the patient was asked to determine whether two sounds played in succession occurred at the same time or not. This was an adaptive staircase task, such that the sounds got closer together each time the subject answered correctly. The sounds started at 5000 ms apart and the delay

between sounds decreased by 1000ms every time the patient correctly answered that the sounds did not occur at the same time, until the delay reached 0ms (i.e., the two sounds converged). At this point, a correct answer (that the sounds were concurrent) resulted in the test repeating from 1000ms, decreasing in steps of 250ms. Thus, the experiment tested the patient's ability to distinguish sounds that were up to 250ms apart, which is approximately the delay at which neurotypical participants are able to distinguish two sounds with no forward masking effect (Jesteadt, Bacon & Lehman, 1982). The test was delivered using Matlab R2013b (Mathworks) with the Psychophysics Toolbox extension (Brainard, 1997). To investigate whether this was a general sound or speech perception deficit, the experiment was repeated three times with three different sets of stimuli: first, two nonspeech sounds (white noise bursts); second, a white noise burst and a recording of a male voice; and third, a non-speech sound and a recording of the patient's speech. Stimuli were approximately three seconds long. This was intended to eliminate the possibility that he had some problem with sound perception that was specific to the acoustic characteristics of speech or of his own voice. The patient successfully completed the task in the minimum number of steps on each of the three trials (that is, he made no errors), and there was no difference between performance on the different sound types. This confirmed that his perception of externally generated sounds and recordings of his own speech was within the normal range.

Preliminary investigations that confirmed that the patient did not experience difficulty hearing and responding to external sounds were followed by an experiment assessing his ability to modulate his voice in response to masking noise. We considered three possible outcomes: if the patient were completely unable to use his self-monitoring system, he might be expected to make no compensation. Conversely if his perception of his own

voice were attenuated relative to external sounds, but he was still able to monitor and make changes to his voice, we might expect over-compensation relative to controls. Finally, if there were no impairment, or if both external and self-produced sounds were attenuated, there would be no difference between his performance and that of controls.

4.5. METHODS

PARTICIPANTS

10 healthy male controls (mean age 51, range 47-56) who reported no speech or hearing disorders were recruited to act as a control group and provided written consent. Their hearing was tested using an Amplivox 116 Screening Audiometer with DD45 earphones (amplivox.ltd.uk). All participants had a four frequency pure tone average hearing threshold of less than 20 dB.

STIMULI

A spontaneous speech task prompted by questions was chosen as the patient had some difficulty reading and describing pictures; additionally, the question-and-answer format added a communicative element to the task, which increased the likelihood of consistent Lombard responses in all subjects. Subjects answered autobiographical questions based on Kopelman, Wilson & Baddeley (1989) while hearing white noise maskers at three different intensity levels. The full list of questions asked is given in Appendix B. Stimuli were created using MATLAB R2013b (Mathworks), and masker intensities were set at 60, 70 and 80 dB SPL using a Bruel & Kjaer artificial ear. These levels were chosen as being within the range that causes vocal adaptation, without causing hearing damage (Cooke & Lu, 2010).

TASK

Subjects heard maskers through Beyerdynamic DT 100 closed-back circumaural headphones and spoke into a RODE NT1-A one-inch cardoid condenser microphone positioned 30cm away from the participant's mouth. Their voices were recorded at 44100Hz with 16 bit quantisation using MATLAB R2013b on a Macbook Pro (Apple), connected to the microphone using a Focusrite Scarlett 2i2 two in/two out USB 2.0 audio interface.

There were 24 trials, each lasting fifteen seconds each; the relatively long trial duration was chosen to give the patient enough time to respond as his speech rate was slowed and he had some difficulty with speech production. Participants sat facing the experimenter, heard the question read aloud and were instructed to press the space bar when they were ready to answer. When they pressed space, the computer screen displayed 'READY', 'SET' and then 'GO' in black text centred on a white background. The noise masker began playing as soon as the 'GO' prompt was shown, and continued for fifteen seconds while the participant responded to the prompt. At the end of the fifteen seconds the noise stopped and the computer displayed a 'STOP' command. The experimenter then read the next question and the participant pressed the space bar when they were ready to proceed. Participants were provided with water and allowed to take short breaks in between trials if they wished. Maskers were randomised across participants using a latin square to control for presentation-order effects, and each trial type was repeated six times. There were additionally five practice trials in which participants experienced each of the different trial types once. The masking experiment lasted for half an hour; the total duration of testing, including the hearing test, was approximately 50 minutes.

ACOUSTIC ANALYSIS

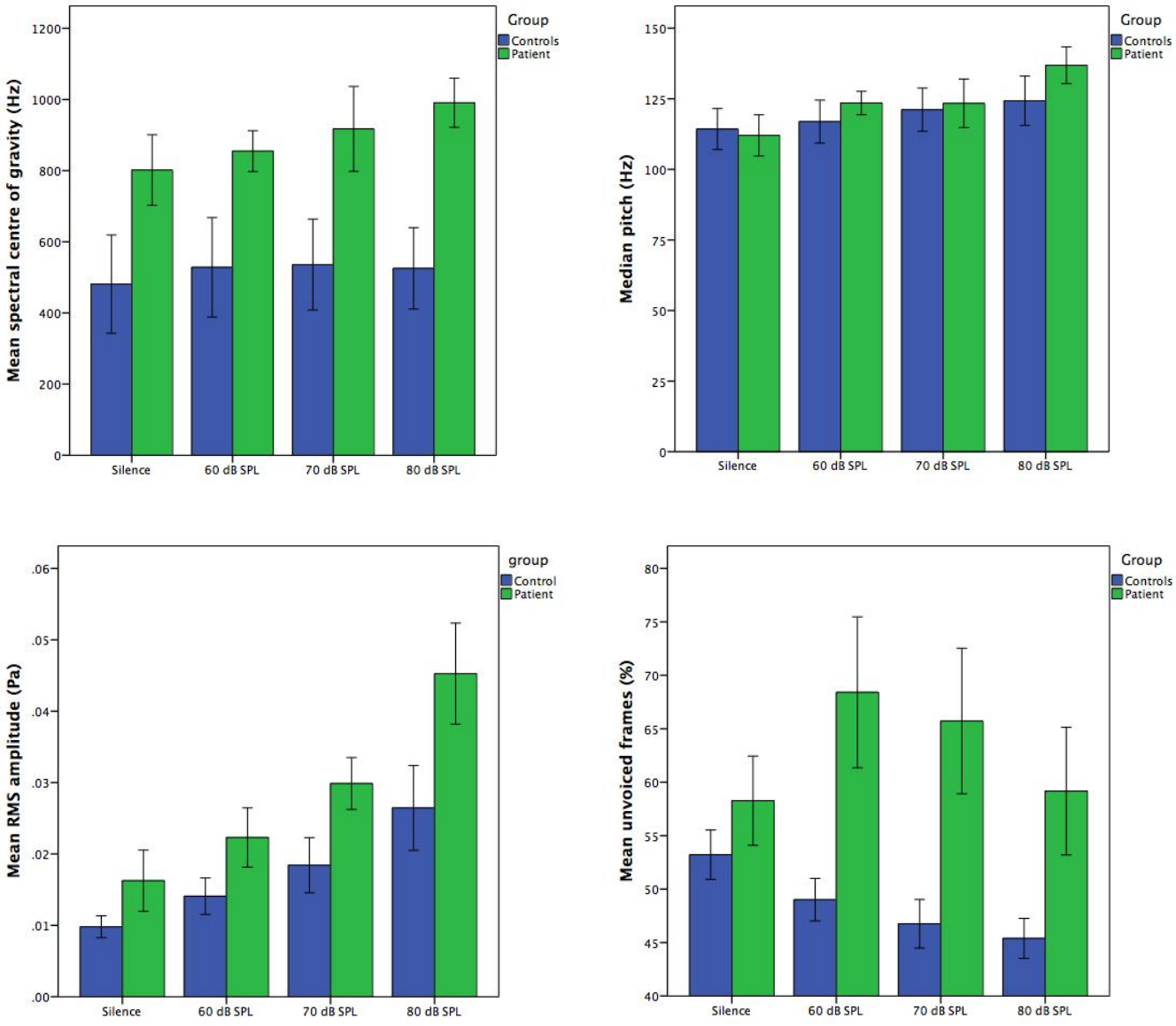
Four acoustic parameters were extracted from the recordings using PRAAT (Boersma & Weenink, 2008): root-mean-square (RMS) amplitude, median pitch, spectral centre of gravity and percentage of unvoiced frames. Median pitch was chosen as it is less vulnerable to outliers, which are especially likely when estimating pitch. Therefore it was considered the most reliable indicator of central tendency for this measure. RMS amplitude results from squaring the amplitude of each point of a waveform, taking the mean of the squared values and calculating its square root. This is a better measure of amplitude than peak amplitude or mean amplitude, as squaring prevents negative values from cancelling out positive ones. Percentage unvoiced frames is a measure of ability to sustain voicing; pathological voices are less able to do this.

Extracted values were analysed using a modified t-test to account for small sample size (Crawford & Garthwaite, 2002). This procedure accounts for small sample size by treating the mean and standard deviation of the control sample as statistics rather than population parameters, and using the t- distribution rather than the normal distribution (which can overestimate abnormality of a patient score, as it has thinner 'tails' than the t-distribution). Additionally, Crawford, Garthwaite and Porter's (2010) modified estimate of effect size, z_{cc}, is given- this is the average difference (in standard deviations) between the patient's score and that of a randomly chosen member of the control population.

4.6. RESULTS

Means and confidence intervals for all acoustic measures in both groups are shown below.

FIGURE 11: ACOUSTIC PROPERTIES OF MASKED SPEECH IN PATIENT VERSUS CONTROLS: MEAN SPECTRAL COG, MEDIAN PITCH, MEAN RMS AMPLITUDE AND MEAN % UNVOICED FRAMES.. ERROR BARS ARE 95% CONFIDENCE INTERVALS, X AXIS IS MASKING CONDITION.



There was a linear correlation between RMS amplitude and masker intensity in both controls (Pearson's r=0.379, p<0.0001) and the patient (r=0.905, p<0.0001). However, there was no significant difference in the strength of correlation between patient and

controls ($t(9)=0.686$, $p=0.51$), nor were there significant differences between patient and control scores on individual conditions. Similarly, median pitch was correlated with masker intensity in both controls ($r=0.184$, $p=0.04$) and the patient ($r=0783$, $p<0.0001$). However, the strength of correlation between did not significantly differ between the patient and controls ($t(9)= 0.291$, $p=0.78$), and there were no significant differences between patient and control scores on individual conditions. Spectral centre of gravity was significantly correlated with masker intensity in the patient ($r=0.669$, $p<0.0001$), but not controls ($r=0.020$, $p=0.76$). However, the modified t-test found no significant difference in the strength of correlation between patient and controls ($t(9)= 0.546$, $p=0.59$), or between patient and control scores in each condition.

There was a linear correlation between percentage of unvoiced frames and masker intensity in controls ($r=-0.381$, $p<0.0001$), in which the percentage of unvoiced frames decreased as masker intensity increased. There was no such linear relationship in the patient data ($r<0.001$, $p=1.00$); rather, there was a nonlinear relationship, such that the percentage of unvoiced frames increased in noise compared to quiet, then decreased with increasing noise levels. There was also a significant difference between the strength of the correlations between patients and controls ($t(9)=2.346$, $p=0.042$) and significant differences between patient and control scores in the 60 dB SPL condition ($t=2.689$, $p=0.02$) and the 70 dB SPL condition ($t=2.466$, $p=0.03$). Additionally, there was no significant difference between patient and control scores in the 80 dB SPL condition ($t=2.183$, $p=0.056$).

TABLE 5: DIFFERENCES IN MEAN ACOUSTIC VALUES BETWEEN CONTROLS AND CASE STUDY

Measure	Masker (dB)	Controls		Case		t	2 tailed p	effect size (z_{cc})	effect size CI	est % of normal population falling below case score	95% CIs for percentage
		Mean	SD	Mean	SD						
Root-mean-square amplitude	0	0.0098	0.0059	0.0163	0.0046	1.029	0.33032	1.079	0.271-1.852	83.4837	60.67-9./80
	60	0.0141	0.0099	0.0223	0.0039	0.792	0.448	0.831	0.089-1.541	77.576	53.53-93.83
	70	0.0184	0.0149	0.0299	0.0034	0.731	0.48	0.767	0.040-1.462	75.84	51.59-92.82
	80	0.02645	0.023	0.0453	0.0067	0.778	0.46	0.816	0.077-1.523	77.18	53.09-93.61
Median pitch	0	112.73	23.63	112.06	6.97	-0.024	0.98	-0.025	-0.645-0.595	49.06	25.96-72.41
	60	116.2	24.78	123.51	3.94	0.281	0.78	0.295	-0.347-0.921	60.76	26.43-82.16
	70	120.73	25.34	123.41	8.2	0.101	0.92	0.106	-0.519-0.725	53.91	30.20-76.56
	80	125.38	28.17	136.88	6.22	0.393	0.7	0.412	-0.246-1.049	64.81	40.28-85.29
Spectral centre of gravity	0	617.72	513.63	801.92	94.43	0.342	0.74	0.359	-0.291-0.99	62.99	38.53-83.90
	60	693.76	585.12	854.98	54.7	0.263	0.8	0.276	-0.364-0.901	60.07	35.79-81.62
	70	667.44	487.74	917.58	113.83	0.489	0.64	0.513	-0.162-1.163	68.17	43.58-87.76
	80	656.97	474.03	990.85	65.75	0.672	0.52	0.704	-0.009-1.387	74.1	49.66-91.73
% unvoiced syllables	0	53.22	7.62	58.27	3.98	0.632	0.54	0.663	-0.041-1.337	72.84	48.36-90.95
	60	49.41	6.74	68.41	6.73	2.689	0.02	2.82	1.386-4.230	98.76	91.71-99.99
	70	47.06	7.22	65.73	6.49	2.466	0.03	2.586	1.245-3.899	98.21	89.34-99.99
	80	45.55	5.95	59.17	5.69	2.183	0.056	2.289	1.064-3.484	97.15	85.63-99.98

4.7. DISCUSSION

Although the perceptual problem reported here has not previously been discussed in the literature, many aphasic patients show an impaired ability to monitor and repair errors in their own speech (Wepman, 1958). Puzzlingly, this deficit occurs even in patients with relatively intact comprehension abilities (Schlenck et al, 1987) and seems to apply only to on-line monitoring, since patients are able to detect errors in a recording of their own voice (Maher, Rothi & Heilman, 1994). We hypothesised that this patient's difficulties perceiving his own voice, but not other sounds (self-produced or otherwise), could be a manifestation of a similar problem. Previous research suggests that patients with expressive aphasia do not change their speech behaviour in noise, and therefore do not rely on postarticulatory monitoring (Oomen, Postma & Kolk, 2001). However, this contrasts with research suggesting that patients with expressive aphasia are more susceptible to the disruptive effect of delayed auditory feedback, which is presumed to result from disruption to the postarticulatory monitoring process. Looking at the acoustic characteristics of this patient's speech in noise, rather than semantic or phonemic errors, provided clear evidence that he compensated for altered feedback. Indeed, these adaptations were consistently at the top of the normal range, although because of variability in the control group this was not always a statistically significant difference. Additionally, while controls' speech showed a decrease in the number of unvoiced frames when talking in noise compared to quiet, in the patient's speech the percentage of unvoiced segments increased as masking level increased. These results could indicate overcompensation and increased effort related to the patient's difficulties with own-voice perception.

One possible reason for the apparent disparity between these results and previous research demonstrating that people with expressive aphasia do not change their speech in noise is that the two studies used masking noise in different ways. In Oomen et al's (2001) study, masking noise was assumed to eliminate the sound of the talker's voice, thus removing the possibility of using auditory feedback to monitor their utterance. Here, however, we accepted that masking noise is unlikely to completely eliminate auditory feedback, as talkers also receive feedback through bone conduction, and any compensation due to the Lombard effect may improve signal-to-noise ratio. Instead, we treated it as a means to attenuate auditory feedback and investigated whether participants changed their voices in response. Additionally, our study and Oomen et al's (2001) defined 'error' differently. For Oomen et al, 'errors' are defined at the word or phoneme level as semantic or phonological slips. Here, the 'error' was introduced by the masking noise and affected the whole utterance.

It seems, then, that there are two groups of findings. The first, from studies that measured corrections of phonetic and syntactic errors, has concluded that patients with expressive aphasia do not rely on feedback monitoring. The second, which looks at the response of patients with expressive aphasia to perturbations of feedback, has concluded that patients with expressive aphasia may over-rely on feedback monitoring. To reconcile these two conclusions, we suggest that attenuated self-perception may drive under-correction at the word and phoneme level, but over-correction when an external source affects the audibility of their voice. Alternatively, there may be different mechanisms for error monitoring at phoneme level and at utterance level, such that one circuit may be impaired while the other continues to work. A future project with this data could involve

analysing the substance of the patient's speech for phonological and semantic errors, so that both types of 'error correction' can be directly contrasted.

CHAPTER 5: MASKED SPEECH PRODUCTION IN TYPICAL SPEAKERS

5.1. ABSTRACT

The study described in Chapter 4 used white masking noise to perturb feedback. This is in keeping with other studies of speech production in masking sounds, such as those described in Chapter 3. These have framed the problem of speaking in the presence of a masker as one of impaired auditory feedback: that is, noise "masks" speech, causing a mismatch between feedback and auditory targets. Increased activity in the superior temporal gyrus (STG), found when speaking in noise compared to quiet, has been interpreted as encoding this mismatch. However, background noise is often a source of information in its own right. This study used sparse fMRI to investigate the contribution of energetic and informational content to neural responses to speaking in a masker.

Participants read sentences aloud in the presence of four different masking conditions, varying in both informational and energetic content—clear speech, rotated speech, speech modulated noise, and continuous white noise. There were three baselines- speaking in quiet, listening to noise, and silent reading.

Analysis revealed increased activity in STG when speaking in noise, compared with speech in quiet. If this resulted from a feedback mismatch, the strongest response should have been to speaking in white noise (the most effective energetic masker). Instead, activation increased with the amount of informational content, with speaking over a competing talker eliciting the greatest response. This pattern remained even when the effect of hearing the different maskers was factored out.

5.2. INTRODUCTION

This chapter describes a study that aimed to develop previous work on speech production in noise by integrating neural and behavioural evidence about the effect of different types of masking sound on speech production. Below, a review of the relevant literature outlines current neural and behavioural research into masked speech production, as well as relevant speech perception research, before describing why it is important to look at a wide variety of masking sounds, rather than simply focussing on continuous broadband noise, as much previous research has done.

NEURAL STUDIES OF AUDITORY FEEDBACK USING MASKED SPEECH

Masking noise is an under-exploited tool for investigating auditory feedback. Most studies using altered feedback have resorted to manipulating the frequency of the talker's voice. Although this technique is widely used and has the advantage of allowing the experimenter to adjust auditory feedback at the phoneme level, if required, the manipulation is of questionable ecological validity; in real life, talkers rarely experience a sudden voice pitch change, unless they have just inhaled helium. Moreover, since the behavioural compensation to a frequency shift occurs over relatively slow latencies, it is often necessary for experimenters to require participants to deliberately prolong their utterances. For example, in one typical frequency-altered feedback study (Tourville et al., 2008) subjects were required to vocalize CVC words for up to 593ms- much longer than they would typically pronounce a vowel in normal speech- and adaptation to altered feedback took around 130ms to occur, by which time the talker would already have moved on to the next phoneme if they were talking at a normal rate (Osser & Peng,

1964b). This means it is difficult to draw conclusions about the effect of altered feedback on normal connected speech at a typical speech rate. By contrast, masking noise allows researchers to investigate how people cope with situations where feedback of their own voice is attenuated in a real-world environment, when speaking at a normal rate. Although so far only a handful of neuroimaging studies have taken advantage of this, in general the results are consistent with studies using other types of altered feedback. Christoffels et al. (2007) compared speech in quiet to speech masked by pink noise of a variable intensity such that it subjectively eliminated the participants' perception of their own voice. Two listening conditions served as a baseline- participants viewed scrambled pictures whilst listening either to a recording of their own voice or to the pink noise masker. Comparing speech in quiet to speech in noise revealed that masked speech resulted in greater activity in bilateral STG, in keeping with other studies of altered feedback. In an attempt to exclude the effects of auditory input, Christoffels et al. (2007) performed a conjunction null analysis of the contrasts SpeakQuiet>SpeakNoise and SpeakQuiet>ListenVoice, which confirmed the results of the original contrast. This conjunction was intended to focus only on activation attributable to receiving accurate feedback- not to hearing sound in general. However, given that speaking in quiet and in noise activates the same areas of cortex as listening, this conjunction essentially limits the analysis to auditory cortex without telling us whether effects within auditory cortex are attributable to either the hearing or the speaking aspect of the task. To make such conclusions it is necessary to compare speaking in noise directly with listening. That said, the results seem consistent with studies that have directly contrasted speech in noise with listening to noise. Zheng et al. (2010), for example, found a significant interaction

between speech production (with and without noise) and listening (to a masker and to a recording of the subject's own voice). Bilateral posterior STG activated for noise conditions compared to speech more in production tasks than when listening to the same sounds. A later re-analysis of the Christoffels et al. (2007) by the same group (van de Ven, Esposito, & Christoffels, 2009) used independent component analysis to try to identify components that activated differently for speech production compared to listening. This analysis revealed one temporal cluster that the authors described as displaying a speech monitoring effect, with listening conditions activating the component more strongly than speaking in noise, which in turn evoked a stronger response than speaking in quiet. Activation in bilateral Heschl's sulcus and parietal areas was strongly related to the component, while activity in SMA was inversely related to it. A follow-up study (Christoffels et al., 2011) investigated the effects of varying noise levels on neural activation while speaking. This found that when speaking in pink noise, activation in right and left STG increases as the level of masking intensity increases, but the same is not true for passive listening to equivalent intensity signals.

BEYOND WHITE NOISE- WHY STUDY SPEECH MASKERS?

The studies that we have looked at to date have primarily investigated the consequences of speaking in white noise. In practice, talkers are as likely- if not more likely- to be communicating in the presence of competing speakers than continuous white noise. To understand why there may be important differences in the way that speech acts as a masker compared to other sounds, it is necessary to recall the difference between energetic and informational masking. Energetic masking results from competition between the target sound and the masker causing overlapping excitation patterns at the

auditory periphery over time. The energetic masking potential of a sound is therefore primarily determined by its intensity and frequency relative to the signal and how the two signals (target and masker) overlap in time. Informational masking, on the other hand, arises from higher-order properties of the signal that cause central competition for resources. For example, when speech is masked by competing speech, informational masking occurs because of the linguistic content of the masker. Although informational maskers do not necessarily need to contain semantic or linguistic content (a police siren may also carry meaning, for example), speech is a particularly interesting example of a signal that is not a very effective energetic masker due to the modulations in its amplitude envelope (Festen & Plomp, 1990), yet is nonetheless an effective masker because it is so high in informational content. It is important to note that all informational maskers necessarily also have a frequency and intensity component that gives rise to energetic masking, but the effects of the two different types of masking are nevertheless clearly dissociable. In speech perception, the intelligibility of speech masked by speech does not show the same monotonic relationship with the signal-to-noise ratio that would be expected if intelligibility was purely a function of the masker's energetic component (Brungart, 2001). In addition, Brungart (2001) found that intelligibility was greater for speech masked by modulated noise with the same temporal and spectral profile as speech, compared to speech masked by speech, even though the two maskers have similar energetic masking potential. That is, the presence of informational masking content makes it more difficult to understand the target signal, even when energetic masking potential is controlled for. There is also an effect of gender beyond that expected from the fact that same-sex maskers are acoustically more similar to the target stimulus

than different sex maskers. Festen and Plomp (1990) compared speech perception in speech-shaped noise that matched the long-term root-mean-square spectra of male and female voices with speech perception masked by same- and different-sex talkers, finding larger differences in performance between the two gendered speech maskers than the gendered speech-shaped noise maskers.

The low energetic masking potential of speech owing to its amplitude fluctuations creates glimpses of the target that listeners may be able to exploit: Cooke (2006) found that the proportion of the time-frequency plane available to listeners through these glimpses is a good predictor of intelligibility. Although in speech production talkers are producing rather than observing the target signal being masked, there may be an analogous effect of type of masker on speech production. For example, cottontop tamarins retime their calls when vocalizing over a patterned noise, in order to exploit gaps in the masker (Egnor, Wickelgren, & Hauser, 2007). Given that the Lombard effect has a communicative component, it is possible that talkers may try to optimise their intelligibility by using the informational content of speech to predict glimpses and retime their utterances to take advantage of glimpses caused by spectral dips and amplitude modulations. Research on this point is limited but suggests that such a strategy is possible. Lu and Cooke (2008) asked subjects to read sentences aloud in six types of noise with varying proportions of energetic and informational masking potential presented at 89 dB SPL. The stimuli consisted of N-talker babble composed of the utterances of one, two, four, eight and sixteen talkers, and speech shaped noise. The one-talker and the SSN condition were also tested at 82 and 98 dB SPL. They found increases in utterance duration, RMS energy, mean F0 and spectral centre of gravity (CoG) consistent with other studies for all types

of noise. At phoneme level, increased N led to increased duration for all phonemes except for /f/ and non-alveolar plosives, for which they observed a slight shortening. There was increased spectral CoG for all phonemes, and flatter vowel spectral tilt. However, the competing talker condition led to smaller utterance-level speech modifications than speech-shaped noise, with modifications increasing with higher numbers of N. This seemed to indicate that changes were largely a function of the energetic masking potential of the noise. The only differences between competing talker and SSN were that effects of spectral CoG and duration of short pauses increased with level for the competing talker only, whereas the voiced/unvoiced ratio increased only for speech-shaped noise. There were differences in short pause duration for competing speech, but no evidence of talkers retiming utterances to exploit glimpses. Lu and Cooke (2008) theorised that this might be because of the lack of a communicative element to the task. To investigate this, they followed up with a study in which talkers solved Sudoku puzzles, alone or in pairs (Cooke & Lu, 2010). This study found that talkers could reduce temporal overlap with the noise, implying that they actively monitored the background and predicted upcoming pauses. Moreover, subjects were better able to retime their speech to exploit spectral and temporal glimpses in a masker when that masker is intelligible speech, as opposed to speech shaped noise. This implies that subjects were able to monitor the environment they were speaking in and use the information from that to optimize their communicative signal. Subsequent studies by the same group have confirmed that talkers retime their voices to reduce overlap with fluctuating maskers (Aubanel & Cooke, 2013; Aubanel, Cooke, & Foster, 2013), although they have not succeeded in confirming a difference between talkers' behaviour in intelligible versus

unintelligible maskers, suggesting that talkers are only able to use semantic information to help retime their utterances in specific, highly communicative settings.

So far, neural studies of speech production in noise have not looked at the difference between informational and energetic masking. However, there is evidence from speech perception that may help to indicate what we may find. In a PET study, Scott, Rosen, Wickham, and Wise (2004) presented listeners with speech masked by a competing talker and speech masked by continuous speech-shaped noise, at a variety of signal to noise ratios. They found that, regardless of SNR, speech perception in steady-state noise was associated with increases in activity in left frontal and prefrontal cortex, and right posterior parietal cortex (when compared to speech masked by speech). By contrast, the intelligible speech masker was associated with the activation of bilateral superior temporal gyri and sulci, extending into Heschl's gyrus in the right hemisphere. Better behavioural performance (i.e. increased target intelligibility) was associated with activation in anterior STG. However, as this study contrasted a continuous masker with the speech masker, differences in activation could be attributed to the fact that the speech masker allowed glimpses of the target signal, while the continuous masker did not. A follow-up PET study (Scott et al., 2009) addressed this by comparing a speech masker with speech modulated noise (which contains glimpses) and rotated speech (which contains glimpses and has a similar harmonic structure to speech). Compared to speech modulated noise, the intelligible speech masker was associated with bilateral STG activation, while rotated speech was associated with STG activation in the right hemisphere only. However, when the speech masker and rotated speech masker were contrasted directly, no significant activity was found; this may reflect a lack of sensitivity

in the imaging technique, or a lack of power owing to the small sample size (eight participants). Finally, Evans, McGettigan, Agnew, Rosen, & Scott (2016)used fMRI to compare neural responses to speech masked by speech, rotated speech and speech modulated noise (SMN), as well as including an unmasked speech condition. The unmasked speech condition was associated with widespread activation in bilateral STG. A smaller subset of this area, in bilateral mid to posterior STG, showed increased activation in line with the informational content of each masker. This implies that informational maskers are processed similarly, but not equivalently, to attended speech. Taken together, the results of these three studies indicate that unattended informational content is processed bilaterally in the STG during speech perception, within the pathway for target speech; it is possible that this also takes place during speech production.

NEURAL EVIDENCE SHOULD BE SUPPORTED BY BEHAVIOURAL DATA

There is a strong justification for looking at how we produce speech in the presence of voices, not just unintelligible noise. Not only is this a situation that talkers frequently encounter in life, but there are important differences in the way that speech acts as a masker compared to other sounds. Previous neuroimaging studies that used speech production in noise to investigate auditory feedback have focused only on the effects of continuous maskers on speech production, and in general have only considered neural activation without analysing any behavioural responses to masked speech production. However, when looking at how people respond neurally to a real-world situation it is important to take on board evidence about how they respond behaviourally to that situation. We have already established that when people with typical hearing speak in noise, they automatically make several acoustic changes known collectively as the

Lombard response (Lombard, 1911) the most noticeable of which is an increase in vocal intensity. Raising your voice increases the signal-to-noise ratio and thus increases the quality of the feedback you are receiving. Eliades and Wang (2012) found that, in macaques, neurons that were more active when vocalizing in noise changed their firing pattern back in the direction of speech in quiet when the macaque exhibited the Lombard response. This, therefore is a potential confound for Zheng et al. (2010), who did not record their participants' voices. Christoffels et al (2007) attempted to address this problem by asking participants not to raise their voices. But the Lombard response is partly automatic, and hard to prevent (Pick et al., 1989). Even though Christoffels et al (2007) report that participants were successful in maintaining a constant vocal level, it is likely that the Lombard response was costly to suppress, potentially confounding the results. That is, any neural activation seen may partially result from the cognitive effort associated with suppressing the Lombard response, rather than from their response to altered feedback. We believe that the most effective way of dealing with this is to incorporate behavioural data into the model rather than ignore or attempt to suppress it. In this experiment, behavioural data was collected alongside fMRI data by recording participants' voices as they spoke during the experiment. Information about talker's vocal intensity extracted from the recording was then entered into the analysis model as a parametric modulator. Although intensity was the only parameter factored into the fMRI model, as it was considered the most reliable correlate of Lombard speech, the study also includes a detailed acoustic analysis of the speech signal produced in the scanner. In addition to intensity, the spectral centre of gravity, spectral standard deviation, harmonic-to-noise ratio (HNR) and utterance duration were measured and analysed.

Spectral standard deviation, or dispersion, measures whether the energy is concentrated mainly around the centre of gravity, or spread out over a range of frequencies. The spectral centre of gravity is the frequency which divides the spectrum into two, such that the amount of energy in both parts is equal. Previous studies (Lu & Cooke, 2008; Varadarajan & Hansen, 2006) have found that Lombard speech is characterized by an energy shift to higher frequencies, meaning that in this study we would expect to see a higher CoG in speech produced in masking noise compared to speech in quiet. Increases in HNR are associated with a perceptually 'clear' voice (Warhurst, Madill, McCabe, Heard, & Yiu, 2012), so may reflect communicative effort. Finally, talkers sometimes exhibit a slower duration or speech rate in Lombard speech (Aubanel & Cooke, 2013; Pittman & Wiley, 2001- but cf Varadarajan & Hansen, 2006), and have likewise been found to slow their speech rate in studies of clear speech produced to counter adverse listening conditions (Picheny, Durlach, & Braida, 1986). Looking at a range of acoustic parameters allows us to assess the effectiveness of the experiment, in addition to informing our interpretation of the neural results.

LOOKING AT DIFFERENT TYPES OF MASKER ENABLES US TO TEST TWO OPPOSING HYPOTHESES

This study sought to integrate neural and behavioural evidence about what talkers do during speech production in noise. Previous neural studies of masked speech production treat the effect of the masker principally as one of attenuated feedback- the louder the noise, the less able you are to hear your own voice and extract information from it. But noise itself can be a source of information—and behavioural evidence suggests that it is one that even non-human primates can exploit. Talkers speak in the presence of competing speech on a daily basis, and behavioural evidence suggests that they may be

able to adopt temporal modification strategies which mitigate the effect of fluctuating maskers on speech communication. Looking at the neural response to a variety of informational maskers may help us better understand this behaviour. However, the value of studying how we speak in different types of masker is not just that it gives us a more nuanced view of an everyday communicative problem. It also provides a test for models of speech production. The way that such models are currently framed suggests that activation in superior temporal cortex is determined by the acoustic similarity between what you hear and what you intended to produce. The amount of "error" in the feedback, thus defined, depends on how well the masking noise occludes feedback—its energetic masking potential. The greater the energetic masking potential, the greater the activation. However, if, as behavioural evidence suggests, talkers actively use the informational content of maskers to modulate their voice, then we might expect the reverse pattern—the greater the informational masking potential, the greater the activation. Here, we aimed to test these two conflicting predictions by analysing neural responses to a range of maskers that varied in their informational content and similarity to speech, investigating—for the first time— neural responses to the challenge of speaking in varied acoustic environments.

5.3. METHODS

MASKER CHOICE AND CREATION

Participants were presented with four maskers: white noise (WH), speech modulated noise (SMN), spectrally rotated speech (ROT), and natural speech (SP). These were intended to represent points on a continuum from strongly energetic, weakly informational masking to strongly informational, weakly energetic masking, with white noise at one extreme and intelligible speech at the other. As white noise has equal energy across the band of audible sound frequencies, it is an extremely effective energetic masker, but contains very little informational content and shares neither the spectral nor the amplitude profile of speech. SMN sounds like a rhythmic rustling noise. It has a relatively constant spectrum equal to the average long term spectrum of the speech stimuli, and shares other features of speech such as amplitude "dips" which allow opportunities to glimpse target sounds when presented as a masker; it is thus a less effective energetic masker than white noise (Cooke, 2006). It is also relatively low in informational content: whilst amplitude modulations may provide participants with some phonemic cues given sufficient context (Bashford, Warren, & Brown, 1996), SMN does not have a harmonic structure or contain any semantic information, and participants did not identify any informational content during the experiment. Rotated speech is a poorer energetic masker than SMN as it contains spectral and amplitude modulations. However, it retains the spectral and amplitude modulations of the original speech signal and is intelligible (Blesser, 1972; 1969), though only with extensive training (which participants in this experiment were not given). To the untrained ear,

rotated speech lacks any semantic content but generates a sense of pitch is similar to speech in its other acoustic properties, with some phonetic features and a quasi-harmonic structure. Finally, intelligible speech has high informational masking potential (including semantic and syntactic information) but contains spectral and amplitude modulations that render it a poor energetic masker. These maskers are not intended to represent equal steps along the scale from high to low energy/information: the difference in energetic masking potential between white noise and speech modulated noise is likely to be much greater than that between speech modulated noise and rotated speech, and rotated speech and speech have theoretically identical energetic masking potential. Rather, the intention was to covary the energetic and informational properties of the four sounds, such that generally, the greater the sound's energetic masking potential, the lower its informational masking potential, and vice versa.

All masking stimuli with the exception of white noise were derived from 20 digital recordings (sampled originally at 22.05 kHz with 16-bit quantization) of the Bamford-Kowal-Bench (BKB) sentence lists (Bench, Kowal, & Bamford, 1979) from a male and female British English speaker. These sentences were chosen as they contained simple vocabulary and syntax making it easier for talkers to comprehend and produce these sentences in the interval between brain acquisitions in the scanner. The BKB sentence lists consist of short sentences (maximum seven syllables) based on utterances from a language sample produced by young hearing-impaired children. The sentences are reasonably consistent in structure and complexity, with phrase structure constrained to the ten most commonly used structures in the language sample, and similar restrictions for morphology and vocabulary (Bench, Kowal, & Bamford, 1979). We included both male

and female speakers to control for a possible gender effect, since in speech perception, same-gender maskers are more effective than opposite-gender maskers (Festen and Plomp, 1990).

Speech modulated noise (SMN) stimuli were derived by modulating a speech shaped noise with envelopes extracted from the original wide-band masker speech signal by second-order Butterworth low-pass filtering at 20 Hz and full-wave rectification. The SMN was given the same long term average spectrum (LTAS) as the original speech. Spectral analysis of the speech signal was carried out using a fast Fourier transform (FFT) of length 512 sample points (23.22 ms) with windows overlapping by 256 points, giving a value for the LTAS at multiples of 43.1 Hz. This spectrum was then smoothed in the frequency domain with a 27-point Hamming window that was two octaves wide, over the frequency range 50 –7000 Hz. The smoothed spectrum was used to construct an amplitude spectrum for an inverse FFT with component phases randomized with a uniform distribution over the range 0–2π. Next, rotated speech was created by inverting the frequency spectrum around a centre frequency of 2kHz, such that low frequencies became high and high frequencies became low (Blesser, 1972). Because natural and spectrally inverted signals have different long-term spectra, all the stimuli were RMS equalized, and speech-based stimuli were low pass filtered to remove energy above 3.8 KHz in order to equate spectral energy across the conditions. Spectrograms and oscillograms of the maskers are given in the figure below, and examples of the stimuli used in each condition are included on the CD of supplementary material.

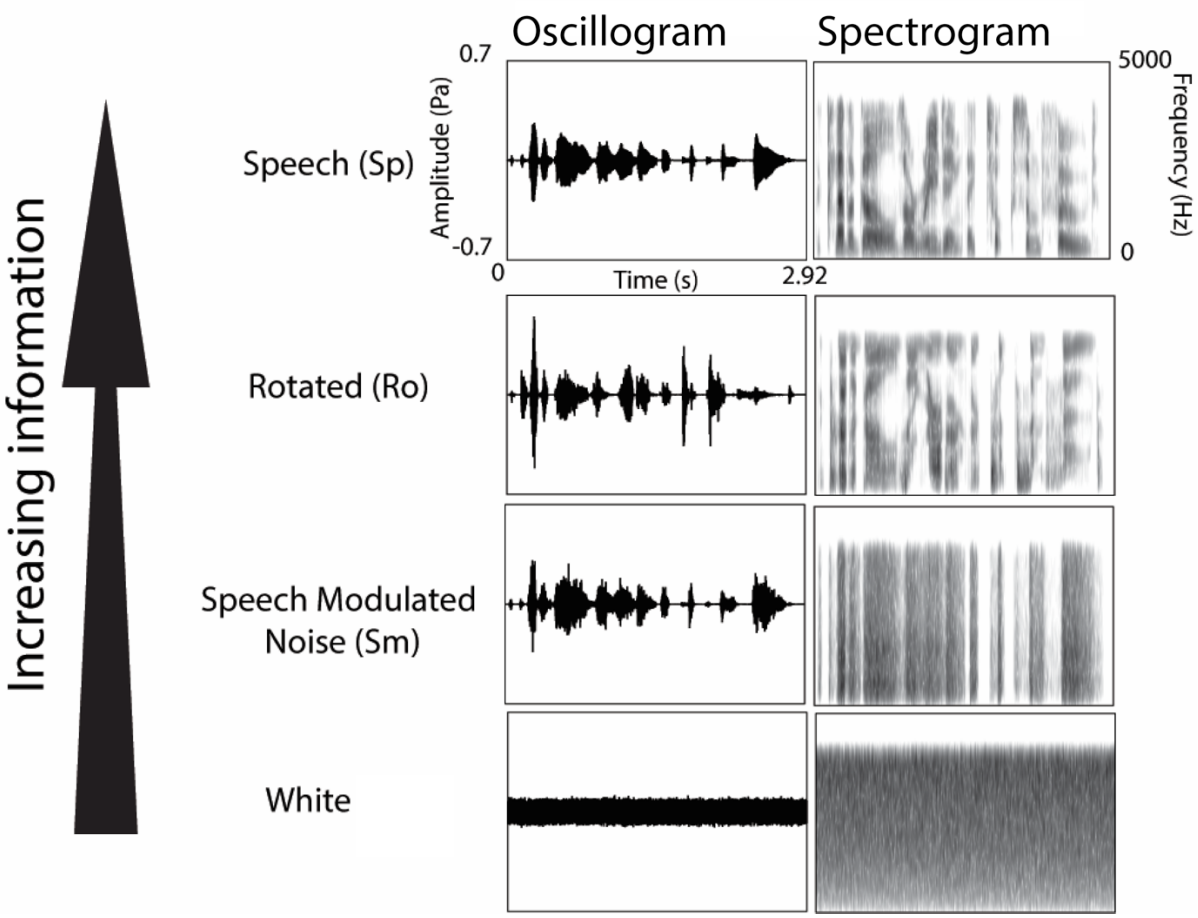


FIGURE 12: OSCILLOGRAMS AND SPECTROGRAMS OF MASKING STIMULI

Each experimental trial consisted of two consecutive BKB sentences (or manipulations thereof) with a silent interval of less than 30ms between sentences. The duration of the white noise and silent trials was fixed to the mean duration of the other maskers (3.2 seconds). Behavioural piloting confirmed that 3.2 seconds was enough time for participants to respond and did not result in long silent periods. While they heard the auditory stimuli, subjects were visually presented with a sentence from the Institute of Hearing Research (IHR) lists (MacLeod & Summerfield, 1990). The IHR sentences are based on the BKB sentence lists with similar syntax, vocabulary and ratio of key words to function words. The words were presented in the middle of the screen in a large and

clearly readable font. To control for higher order processes such as semantic processing involved in reading, participants always saw sentences regardless of whether they were being presented with a masker or not, meaning that the baseline condition was reading silently in quiet.

Behavioural Pre-testing

To ensure that the maskers and noise levels were sufficient to elicit vocal changes, 17 adults (aged 20-37, 8 females) participated in a behavioural version of the experiment. None of the pilot participants took part in the subsequent fMRI study. This study and the fMRI follow-up were approved by the UCL Psychology Research Ethics Committee, and all participants gave written consent.

In the behavioural pilot, participants read IHR sentences aloud while hearing silence, one of the unintelligible maskers (white noise, rotated speech, SMN), male speech or female speech. The natural speech condition was split by gender in order to investigate whether being masked by a speaker of the same gender causes more difficulty or requires different vocal adaptations than talking while being masked by a speaker of the opposite gender- analogous to the increased masking potential of same-sex speech-on-speech masking in perception (Brungart, 2001). All maskers were presented at 84 dB SPL apart from white noise, which was presented at 71 dB SPL, as piloting revealed that participants perceived white noise played at 84 dB SPL to be louder than other maskers presented at the same level. 71 dB SPL was chosen as the level best matched perceptually to the other maskers after piloting several different levels of white noise.

Participants' speech was recorded and analysed for several acoustic changes typically seen in Lombard speech- intensity, utterance duration, median pitch and spectral centre of gravity. In all noise conditions, one-sample t-tests conducted for each acoustic parameter demonstrated that in every condition the features measured increased significantly relative to quiet ($p<0.05$). Within-subjects ANOVAs showed a significant main effect of noise condition for intensity ($F(4,68)= 3.71$, $p=.009$, $\eta_p^2= .179$). The means plot (Fig 13) showed higher intensity in white noise than in the other conditions, but post-hoc tests (corrected using Sidak-adjusted alpha levels) found that the differences between conditions did not reach statistical significance.

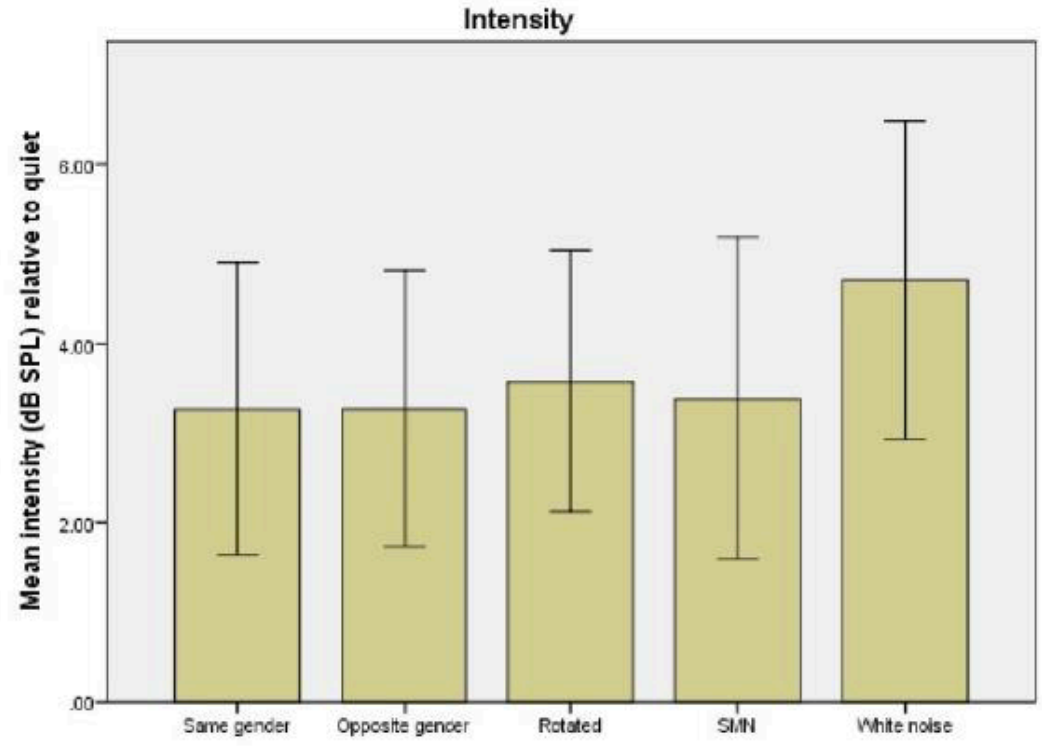


FIGURE 13: MEANS PLOT OF INTENSITY RELATIVE TO QUIET IN DIFFERENT MASKERS (BEHAVIOURAL PRETESTING)

There were no significant differences between conditions for any of the other parameters measured. It was considered that lowering the level of the white noise had reduced its energetic masking potential, contributing to the lack of significance. For the fMRI experiment, therefore, white noise was presented at the same level as the other maskers. As there was no observable difference between the gendered maskers the two were conflated to form one clear speech masking condition for the fMRI experiment.

FMRI SCANNING

PARTICIPANTS

Sixteen right-handed native British English talkers provided written consent (7 females, 9 males; aged 21-38; mean age 29) and were paid a nominal fee for their time. All participants spoke with a Southern British English accent and reported no history of hearing or language impairment. Two participants (one male and one female) did not consistently follow the task instructions (i.e. remained silent when they were meant to speak or spoke when they were meant to listen) and were excluded. Functional and behavioural analyses were conducted on the remaining 14 subjects (6 females, 8 males).

ACQUISITION PARAMETERS

Subjects were scanned on a 1.5T MRI scanner (Siemens Avanto, Siemens Medical Systems, Erlangen, Germany) with a 32- channel head coil. Functional MRI images were acquired using a T2- weighted gradient-echo planar imaging sequence, which covered the whole brain (TR=10s, TA=3s, TE=50ms, flip angle 90 degrees, 35 axial slices, matrix size=64x64x35, 3x3x3mm in-plane resolution). High-resolution anatomical volume images were also acquired for each subject. (HIRes MP-RAGE, 160 sagittal slices, matrix

size: 224x256x160, voxel size=1 mm^3) The field of view was oblique angled away from the eyes (to avoid ghosting artefacts from eye movements) and included the frontal and parietal cortex at the expense of the inferior temporal cortex and inferior cerebellum.

TASK

In the scanner, visual and auditory stimuli were displayed using MATLAB R2013b (Mathworks) with the Psychophysics Toolbox extension (Brainard, 1997). Subjects heard sounds presented through Sensimetrics S14 fMRI-compatible insert earphones, and spoke into an OptoAcoustics FOMRI-III noise-cancelling optical microphone. At the same time, the sentence to be read was projected onto an in-bore screen, using a specially-configured video projector (Eiki International). All the sounds were played at 84 dB SPL as measured by a Bruel & Kjaer 4153 artificial ear outside the scanner on Beyerdynamic DT100 headphones, although it should be noted that it was not possible to confirm the sound level as delivered to the subjects via the Sensimetrics earphones, as sound intensity changes relative to the magnetic field when using these earphones, and MR-safe calibration equipment was not available. However, all participants used the same equipment and their heads were placed in the same position relative to the magnet, so the intensity level remained consistent across subjects. Subjects were trained to perform the experiment outside the scanner on a laptop and were allowed to practise until they were comfortable with the task and were able to respond accurately and quickly.

Participants were trained to read aloud or silently, depending on the colour of the text presented on-screen. If the text was black, they read it silently to themselves; if it was red, they spoke the sentence aloud. At the same time, they heard one of the masking sounds, or silence. This gives us four main experimental tasks: reading silently, hearing nothing

(Rest); reading silently, hearing maskers (Listen); reading aloud while hearing nothing (SpeakQuiet); and reading aloud while hearing maskers (SpeakNoise). The SpeakNoise condition consisted of four separate conditions, one for each of the masking noises: SP, ROT, SMN, and WH. Because of constraints on experiment duration and participants' attention, we made the choice to include one listening condition containing all of the maskers, rather than four separate listening conditions, one for each of the maskers. This means that the Listen task was one condition composed of a combination of sounds from the four masking conditions. This was intended as an approximate control for activation resulting from auditory processing related to hearing the different masking sounds in the SpeakNoise condition

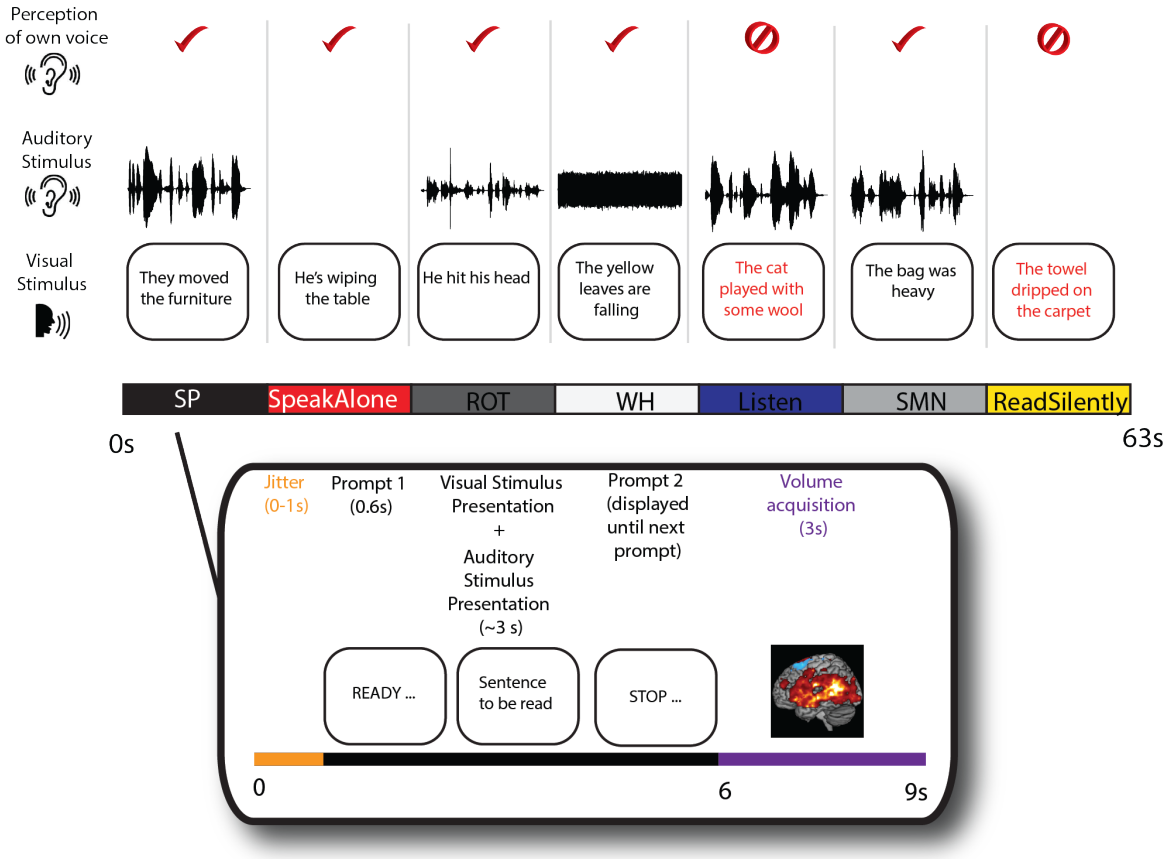


FIGURE 8: PRODUCTION OF SPEECH IN MASKING SOUNDS: EXPERIMENTAL PARADIGM. EACH BLOCK IN THE MIDDLE BAR REPRESENTS ONE TRIAL; INSET SHOWS SEQUENCE AND DURATION OF EVENTS DURING EACH TRIAL.

In SpeakNoise trials, participants spoke for the duration of the masking sound; if they spoke after the noise had finished these trials were excluded from acoustic analysis and were coded as errors in the fMRI design matrix. SpeakQuiet trials were excluded if participants continued to speak for longer than 3.2s (the average trial length for the noise), or if they failed to obey the task instructions (speaking when they were meant to remain silent or being silent when they were meant to speak). These errors occurred very infrequently (mean of number errors per participant =3 of 270 trials, min=0/270, max=10/270) except in the case of two excluded participants. Subjects were told to speak as clearly as possible when reading aloud as someone within the console room would be scoring their speech intelligibility, as heard over the intercom. They were not specifically prompted to speak loudly.

Participants took part in two functional runs, each consisting of 20 trials per condition (SP, ROT, SMN, WH, SpeakQuiet, Listen) and 15 ReadSilently baseline trials, making a total of 135 trials per subject. Every trial consisted of two sounds (or a silent period) lasting about 3.2s on average with one sentence presented on the screen for the subject to read. Masking stimuli were repeated across runs, with the order independently randomized within each run, but the visually presented sentences were all unique. The 15 silent trials were distributed at regular but unpredictable intervals throughout each run, while the other conditions were randomly permuted in sets of six such that each condition was represented once every six trials. This ensured that at most there could be a single consecutive repetition of a condition type.

To ensure that the stimuli were presented in silence and to minimize any susceptibility artefacts caused by the subjects speaking, slow sparse acquisition was used. Each trial

was randomly jittered by 0, 0.5 or 1s. Participants then saw a visual prompt "READY …" which lasted 0.6s, followed by the presentation of a sentence displayed on screen for the participant to read for the duration of the masking sound (or 3.2s in the case of the quiet and listen conditions). A "STOP" prompt was displayed following the offset of the masker and was displayed during the volume acquisition until the subsequent "READY …" prompt.

FMRI PREPROCESSING AND FIRST-LEVEL ANALYSIS

Functional and structural images were analysed using Statistical Parametric Mapping (SPM 8).

To allow for T_1 saturation effects, the first three functional volumes of each run were discarded. Scans were realigned to the first volume by six-parameter rigid-body spatial transformation. The mean functional image was written out and coregistered with the T1 structural image. The estimated translation (x,y,z) and rotation (roll, pitch, yaw) parameters that resulted from motion correction were inspected and did not exceed 3mm or 3 degrees in any direction.

Scans were spatially normalized into MNI space at $2mm^3$ isotropic voxels using the parameters obtained from the unified segmentation of each participant's T1-weighted scan using the ICBM tissue probability maps, and smoothed using a Gaussian kernel of 8 mm^3 at full-width-half-maximum to ameliorate differences in intersubject localization.

First-level analysis was carried out modelling the conditions of interest: Speech in noise: (1) SP (2) ROT (3) SMN (4) WH, (5) SpeakQuiet (QU) and (6) Listen (LI), all with silent trials as an implicit baseline. In addition, first-level contrasts were generated for each of

the speech production conditions (SP, ROT, SMN, WH, QU) with Listen as the baseline. Events were modelled from the coincident presentation of the written text with sound using a canonical hemodynamic response function. This was intended to capture neural processes associated with any delay in speaking following the onset of the masker, which might represent cognitive effort or lexical decision processes. However, on average, the difference between trial onset and speaking onset was very small (0.57 seconds). An analysis modelling events from the onset of speech found equivalent results, indicating that these parameters did not have a significant effect on our analyses.

REGRESSORS

For each condition in which spoken output was required, a parametric regressor modelled variation in RMS amplitude of the speech produced on each trial, measured post hoc using the within scanner recordings. As a proxy for vocal change induced by speaking in noise, this removed neural activity associated with within condition variance in vocal loudness (Wood, Nuerk, Sturm, & Willmes, 2008). Tests for violation of sphericity indicated that this variance was larger in the noise condition than when participants spoke in quiet ($p<0.001$). By modelling out *within* condition variance in neural responses using parametric regressors we hoped to more sensitively identify differences in mean activity *between* conditions. Errors occurred when a participant spoke when they were required to remain silent, remained silent when they were meant to speak, or spoke for longer than the 3.2s recording window. Each error was coded in an additional regressor and the event was removed from the appropriate condition regressor. The model also included six motion parameters of no interest and a Volterra expansion of those parameters (18 regressors in total), shown previously to reduce movement related

artefact (Lund, Nørgaard, Rostrup, Rowe, & Paulson, 2005). In total, therefore, there were 36 additional regressors per run.

Second-level analysis

These contrasts were taken up to a second level random effects model to create two ANOVAs: one looking at the difference between BOLD responses during the three different tasks (SpeakNoise, SpeakQuiet and Listen) with Rest as the baseline, and another looking at differences between responses to speaking in the different masking conditions (SP, ROT, SMN and WH) relative to Listen (as an attempt to control for auditory activation related to just hearing the masker). At the group level, contrasts were thresholded using a voxel wise familywise error rate (FWE) correction for multiple comparisons at $p < 0.05$. Statistical images were rendered on the normalized mean functional image for the group of participants.

5.4. RESULTS

BEHAVIOURAL RESULTS

Audio recordings from the scanner were manually edited to remove silent periods at the start and end of each trial. There was a very quiet repetitive noise in the background from the scanner helium pump, which was filtered out using the method described by Rafii and Pardo (2011). Any residual noise that survived the filter was distributed equally across conditions so should not affect interpretation of the data. The recordings were analysed using Praat (Boersma & Weenink, 2008), and the data extracted was evaluated using IBM SPSS Statistics (version 20).

The following acoustic parameters were extracted: mean intensity (measured in dB relative to the auditory threshold), median F0, spectral centre of gravity, mean harmonic-to-noise ratio (HNR), mean duration and spectral standard deviation.

F0 was computed using the auto-correlation method, with pitch floor set at 75 Hz and pitch ceiling at 1000Hz. Changes in pitch were assessed using the median, as the pitch estimation was less affected by outliers caused by occasional failure to accurately track the pitch of the utterances using the automated pitch tracking algorithm within Praat. Spectral centre of gravity and standard deviation (calculated using the power spectrum) were used to track changes in the distribution of energy across the spectrum. Mean HNR was the mean ratio of quasi-periodic to non-period signal across time segments. Mean duration was evaluated after the sentences had been manually trimmed for silence at the beginning and end of a word.

We used a linear mixed model to investigate the relationship between noise condition and acoustic properties of speech, with condition as a fixed effect, crossed random effects for subjects and sentences read, and a by-subjects random slope for the effects of condition. This was intended to handle the correlated subject data and address the fact that both subjects and sentences are sampled from a larger population (Barr, Levy, Scheepers, & Tily, 2013; Clark, 1973)

This model showed no effect of masking condition on spectral centre of gravity ($F(4,61)=1.51$, $p=.209$), mean HNR ($F(4,53.8)=1.85$, $p=.132$) (or median pitch ($F(4, 2454)=.476$, $p=.754$). A significant effect of masker on mean duration ($F(4,58.4)=2.208$, $p=.016$) was driven by a trend towards increased duration in the masking conditions compared to quiet, but these differences did not survive Sidak correction for multiple comparisons. However, intensity was significantly affected by masking condition ($F(4, 54)=24.15$, $p<0.001$), and Sidak-corrected post-hoc comparisons revealed that intensity was significantly greater in ROT, SM and WH than SP or QU ($p<.001$). There were no significant differences between SP and QU ($p=.989$). There was a statistically significant linear trend ($F(1, 13)=7.85$, $p=.015$, $\eta p2 =.377$) in which intensity increased as the energetic content of the masker increased. There was also a significant effect of masking condition on spectral standard deviation ($F(4,60.17)=3.50$, $p=.012$), caused by a significant decrease in spectral standard deviation in the SM condition compared to SP. There were no other significant differences between conditions.

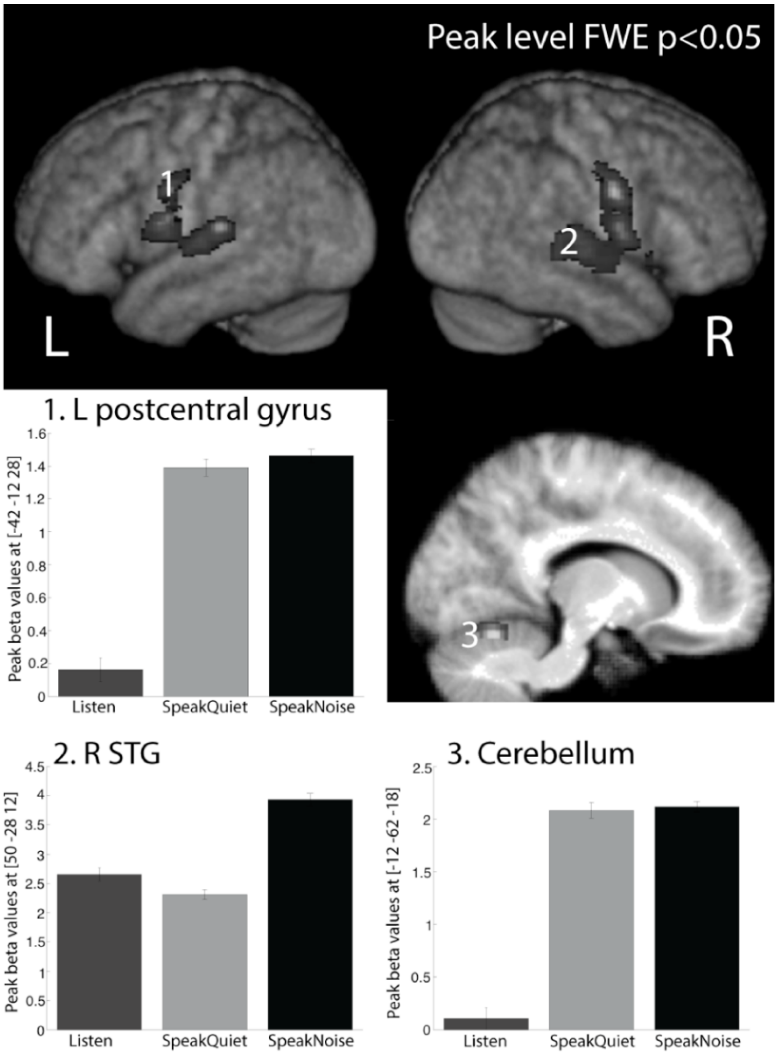


FIGURE 15: BRAIN REGIONS SIGNIFICANTLY MODULATED BY THE THREE DIFFERENT TASKS, THRESHOLDED AT VOXELWISE FWE P<0.05 WITH SILENT READING AS A BASELINE.

FMRI RESULTS

The perception of sounds (speech, rotated speech, SMN and white noise) in the Listen condition was associated with activation of the dorsolateral temporal lobes (including superior temporal gyri). In contrast, speech production (both in silence and in masking sound) was associated with activation in auditory and sensorimotor cortical fields. To look more specifically at the differences between tasks, we conducted an F-test, FWE-

corrected at the whole brain level using a significance threshold of $p<0.05$. This confirmed that activation in the bilateral postcentral gyri was significantly greater in the two speaking conditions than in the Listen condition, with no significant differences between SpeakQuiet and SpeakNoise. In temporal cortex, activation was seen bilaterally in regions covering most of the STG with peaks at [-52 -28 10] and [-60 -30 18] in the left, and [50 -28 12] and [54 -18 8] in the right. Across these regions, the response to the SpeakNoise condition was significantly greater than to SpeakQuiet or Listen.

TABLE 6: PEAK VOXEL CO-ORDINATES REVEALED BY AN ANOVA COMPARING THE THREE TASK CONDITIONS (SPEAKNOISE, SPEAKQUIET AND LISTEN), WITH THE REST CONDITION AS A BASELINE. CORRECTED FOR MULTIPLE COMPARISONS AT FWE P<0.05

Anatomy	Voxels (k)	Z-score	X	y	z
Cerebellum Lobule VI	726	7.36	-12	-62	-18
Cerebellum Lobule VI		7.11	12	-64	-16
Left postcentral gyrus	2747	6.85	-42	-12	28
Left STG		6.65	-52	-28	10
Left STG		6.53	-60	-30	18
Right STG	2751	6.74	50	-28	12
Right postcentral gyrus		6.64	58	-4	36
Right STG		6.23	54	-18	8
	13	5.42	10	-28	-6
Left Insula	27	5.37	-34	8	4
Right Pallidum	57	5.34	28	-4	-6
Right Pallidum		5.18	28	-12	-2
Right Insula	32	5.29	40	12	6
Thalamus- parietal	3	4.96	-12	-26	-4
Right inferior frontal gyrus	8	4.95	54	14	0

We saw a response that could be characterised as speaking-induced suppression in bilateral STG, where speaking in quiet resulted in a reduction of activity relative to passive listening. Although the difference between conditions was only statistically significant in the left hemisphere a comparison of the activation at peak voxels in STG identified by the whole brain analysis using a two-way repeated measures ANOVA revealed no significant effect of hemisphere ($F(1,13)=.188$, $p=.67$, $\eta_p^2=.014$) , or any significant task*hemisphere interaction ($F(2,26)= 2.45$, $p=.106$, $\eta_p^2=.159$), indicating that there was no significant lateralization of brain response to speech in quiet vs. listening at these locations in the STG.

The two speaking tasks (SpeakNoise and SpeakQuiet) were associated with bilateral activation in postcentral gyri and in cerebellar lobule VI. In these regions, responses were significantly greater in the two speaking conditions than in the listening condition, but there were no significant differences between the two speaking conditions, suggesting that this activation reflects a general motor network supporting articulation. Next, to establish modulation of brain activity associated with speaking in the different maskers, we conducted an F-test at the whole brain level (FWE corrected at $p<0.05$) looking at the differences between each of the speech production conditions (SP, ROT, SMN, WH and SpeakQuiet), contrasted with listening as a baseline (Figure 16). This was intended to factor out activation in auditory areas caused by just hearing the masking noise, while

revealing only areas that were associated with the act of speaking in noise.

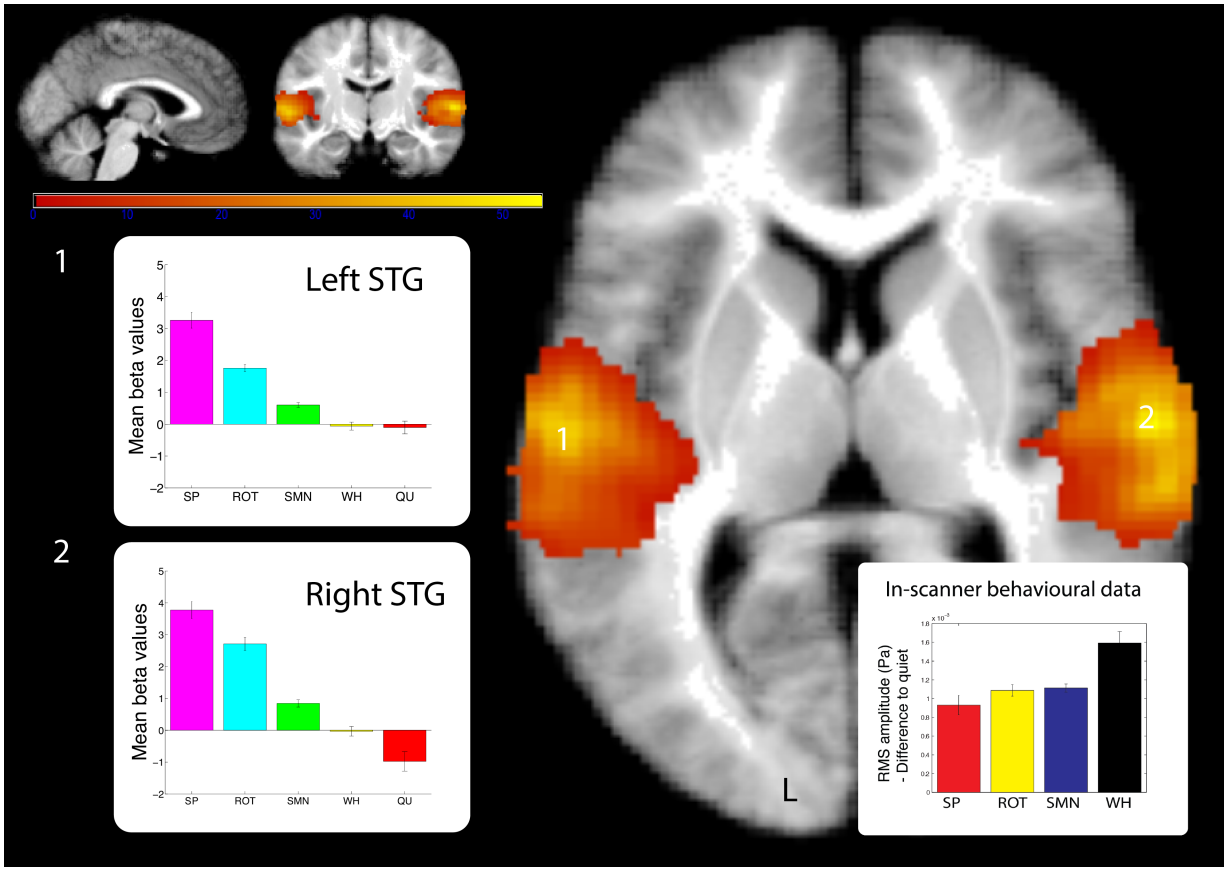


FIGURE 16: EFFECTS OF CONDITION IN BILATERAL SUPERIOR TEMPORAL CORTICES, THRESHOLDED AT FWE P<0.05 WITH LISTEN AS A BASELINE. BAR GRAPHS SHOW MEAN BETA VALUES AT PEAKS [-58 -12 2] IN THE LEFT HEMISPHERE AND [62 12 6] IN THE LEFT HEMISPHERE. INSET: EFFECTS OF MASKING CONDITION ON VOCAL AMPLITUDE, WITH VOCAL AMPLITUDE IN QUIET AS A BASELINE.

TABLE 7: PEAK VOXEL CO-ORDINATES REVEALED BY AN ANOVA COMPARING THE FIVE SPEECH CONDITIONS (QU, SP, RO, SM, WH) WITH THE LISTEN CONDITION AS A BASELINE. CORRECTED FOR MULTIPLE COMPARISONS AT FWE P<0.05

Anatomy	Voxels (k)	Z-score	x	Y	z
Left STG	2302	Inf	-58	-12	2
Left STG		6.56	-44	-30	12
Middle temporal gyrus	6.52	-60	-32	8	
Right STG	2289	Inf	62	-16	6
Right STG		7.77	64	-6	0

Right STG		7.37	52	-24	14
Right STG	7	5.07	50	-46	16

This analysis revealed activation in the left middle temporal gyrus and bilateral superior temporal cortices. In both left and right temporal cortex the response was greatest for speaking over speech, with activation decreasing as the amount of informational content in the masker decreased. At peak [-58 -12 2] in the left STG, a one-way repeated measures ANOVA revealed a significant effect of masking condition ($F(1.5, 19.6)= 61.8$, $p<.001$, $\eta_p^2 =.826$). Sidak-corrected posthoc tests showed that responses in the SpeakQuiet and white noise (WH) conditions were not significantly different from each other ($p= 1.0$), and there was also no significant difference between responses in the SpeakQuiet and SMN conditions ($p=.053$). One-sample t-tests with a test value of 0 (representing the listening baseline) showed that activity in the SpeakQuiet and WH conditions was not significantly different from baseline. All other conditions were significantly different from the baseline and from each other ($p<.05$). In the right hemisphere, at peak [62 -16 6] in the STG, a similar pattern of activation was seen. Neither WH nor SpeakQuiet were significantly different from baseline. However, there was a significant effect of masking ($F(1.6, 20.8) = 63.7$, $p<.001$, $\eta_p^2 =.831$), and Sidak-corrected post-hoc tests confirmed that all conditions were significantly different to each other ($p<.05$).

At the whole brain level we did not see any regions that responded more to energetic masking than to informational content. In order to conduct a more sensitive search for regions that might display this response, we conducted a region of interest (ROI) at peaks in which speaking induced suppression was identified (defined as a reduction in

activation in the SpeakQuiet condition relative to Listen and SpeakNoise). This response profile was considered to identify feedback-sensitive regions which were involved in encoding mismatch, and could therefore be expected to respond more to energetic masking potential. From the task ANOVA two peaks were identified as fitting this profile, one in the left STG at [-52 -28 10] and one in the right STG at [52 -28 10]. A spherical ROI of radius 8mm (the size of the smoothing kernel) was built around each of these points using the MarsBaR toolbox for SPM (Brett, Anton, Valbregue & Poline, 2002). Within each of the two ROIs an ANOVA was carried out to evaluate differences between the SpeakNoise conditions (SP, ROT, SMN, WH) relative to the baseline of silent reading.

In the left STG ROI, one-way repeated measures ANOVAs revealed a significant effect of masking condition ($F(3,39)= 35.424$, $p<0.001$, $\eta_p^2=.732$); Sidak-corrected post-hoc tests showed significant differences between all conditions except for SMN and White. There was a statistically significant linear trend in which greater BOLD responses were seen for maskers with more informational content ($F(1,13)=54.65$, $p<0.001$, $\eta_p^2=.808$). There was also a significant effect of masking condition in the right STG ROI ($F(3,39)= 17.428$, $p<0.001$, $\eta_p^2= .573$). Post-hoc Sidak-corrected t-tests showed that while there were no significant differences between responses to SP and ROT, or between SMN and WH, all other conditions were significantly different from each other ($p<0.05$). There was also a statistically significant linear trend in the data ($F(1,13)=31.194$, $p<0.001$, $\eta_p^2= .706$),

indicating that the BOLD response was greater for maskers with greater informational content.

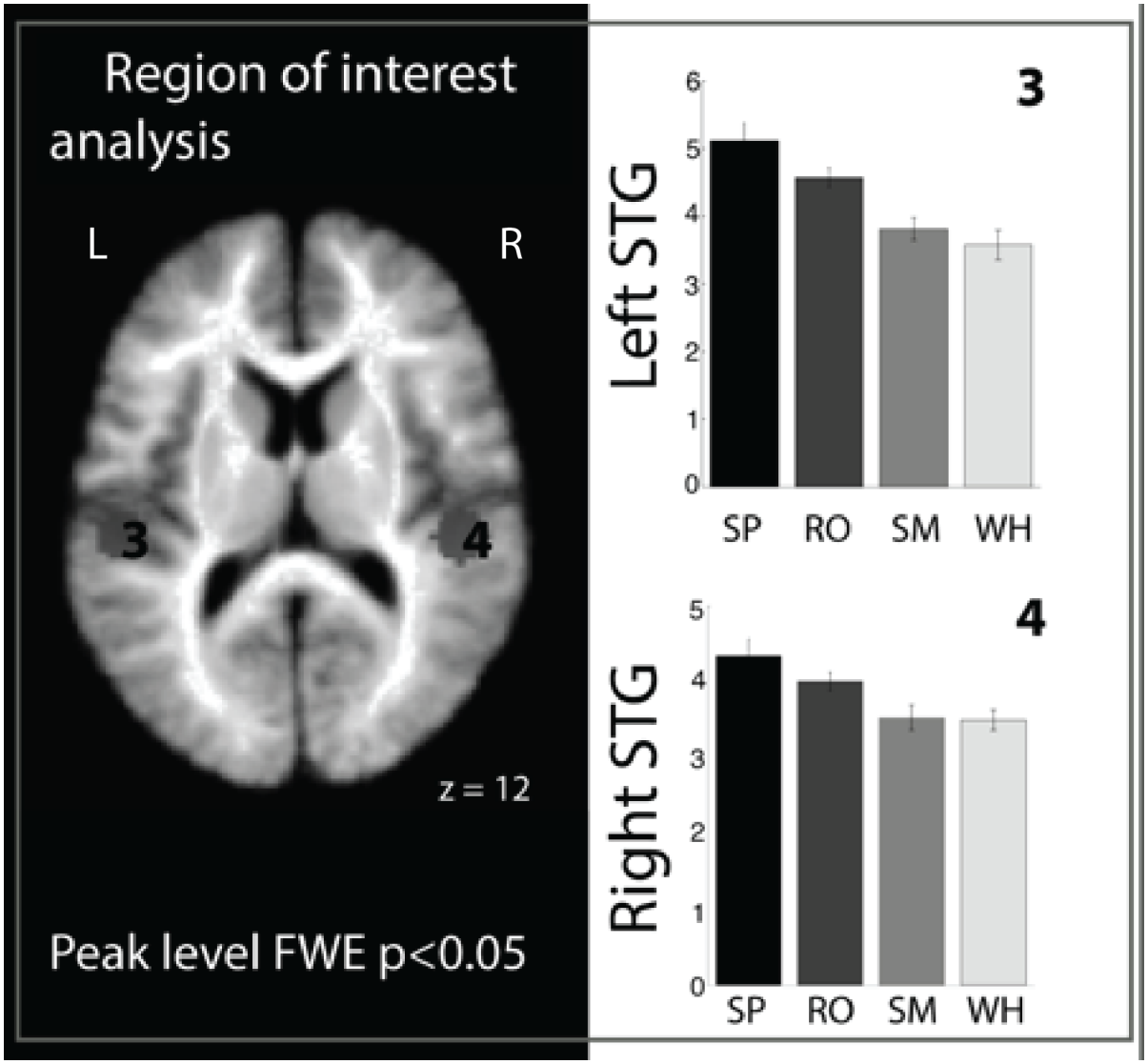


FIGURE 7: MEAN BETA WEIGHTS IN EACH OF THE FOUR SPEAKING CONDITIONS REVEALED BY REPEATED MEASURES ANOVAS LOOKING AT DIFFERENCES BETWEEN THE FOUR MASKING CONDITIONS IN TWO 8MM SPHERICAL REGIONS OF INTEREST CENTRED AROUND [-52 -28 10] IN THE LEFT HEMISPHERE AND [52 -28 10] IN THE RIGHT HEMISPHERE.

In this analysis, we found no neural profiles that correlated with the direction of behavioural vocal modification, i.e. where the greatest response was to talking in continuous noise, and the weakest response was to speaking against another talker. As a final check, we carried out a contrast subtracting activation in the SP (speech-in-speech) condition from activation in the WH (speech-in-noise) condition. This contrast was

designed to test for regions that responded more to speaking in energetic than informational masking; however, the analysis revealed no activation even at a weak threshold of uncorrected $p<0.0005$.

5.6. Discussion

This study aimed to evaluate claims that the superior temporal cortex is an auditory error monitor that activates when what we hear (auditory feedback) does not match up to what we intended to say (auditory target). Conversely, if there is no mismatch between auditory feedback and the intended target, activation will be suppressed. In terms of neural activation, this theory predicts that speaking in quiet (with error-free feedback) will cause a speaking-induced suppression response in superior temporal cortex. This response is characterised by a relatively lower BOLD response in the STG when speaking in quiet, compared to passive listening. Conversely, when speaking in altered feedback, the mismatch between feedback and target results in a release from this suppression, meaning that the BOLD response in STG should be the same as or greater than activation during passive listening. In this experiment, we found a bilateral pattern of STG activation which appeared to reflect just such a speaking-induced suppression response. In keeping with previous neural investigations into speech production in noise, we expected that within this region, the neural response to producing masked speech would correlate with the energetic masking potential of the masker. The logic behind this claim (Christoffels et al., 2007) is that the masker impairs the quality of feedback you receive; so the more effectively the noise prevents you from hearing your voice, the less accurate feedback is, and therefore the greater the 'auditory error'. Thus, in this experiment we would expect

to see STG feedback regions respond most to speaking over a white noise masker- since of the four conditions, this was the most effective energetic masker. Speech was the least effective energetic masker, so we might expect to see the least activation in this condition, with the other maskers somewhere in between. In fact, the opposite pattern was found. Responses to white noise were not significantly greater than listening, and activation increased with informational, not energetic, masking potential.

The relative deactivation in white noise compared to other maskers might be explained by the behavioural data—on average, speakers increased their vocal level most in white noise. This increased amplitude will have improved the signal-to-noise ratio, potentially causing a move back towards the activation patterns seen in quiet, as Eliades and Wang (2012) observed in macaques. Alternatively, although efforts were made to control for the fact that participants were hearing something different in each condition, the fact that there was just one listening control condition in which a combination of all the maskers was played, rather than a different control for each masker, means that differences between conditions may still be the result of auditory activation, rather than speech strategies. Right superior temporal cortex responds to sounds with dynamic pitch (e.g., speech and rotated speech), while left superior temporal cortex responds preferentially to intelligible stimuli (Scott, 2000), which may explain the strong STG response to speech in both hemispheres. However, even with these caveats in mind, it is remarkable that no region of the brain showed a preferential response to energetic over informational masking, even at an uncorrected threshold. The dominant cortical effect of informational masking during speech production seen here suggests that talkers process unattended speech to a high cortical level. The activation seen here in bilateral superior temporal

cortices is similar to that found in studies of speech perception in informational and energetic masking (Scott et al., 2004, 2009; Evans et al, 2016), where masking speech leads to extensive activation in bilateral superior temporal lobes, in addition to the activation seen to attended speech. This strong cortical effect of informational masking may underlie the kind of intrusions from the unattended masking speech that is seen in both speech perception (Brungart & Simpson, 2001) and speech production (Cherry, 1953) paradigms, as well as the more specific ways that speech production can be affected by concurrent masking sounds (Cooke & Lu, 2010).

Behaviourally, we found that speakers reliably increased the RMS amplitude of their voice in noise compared to quiet, and there were also differences between adaptations to different conditions. Notably, several acoustic responses to speaking in noise relative to quiet that have been observed in other studies (Cooke & Lu, 2010; Lu & Cooke, 2008) such as increased spectral centre of gravity and increased pitch, were not seen here. These adaptations were seen in the behavioural pilot, demonstrating that the experimental manipulations were sufficient to induce voice change outside of the scanner, but not within it. This may be because of physiological considerations—the subjects were lying supine in the scanner, which affects vocal tract shape and articulator positions (Kitamura et al., 2005). Alternatively, participants may not have been motivated to maximize their communicative efforts (despite being told they were being scored for intelligibility) because they were vocalizing on their own in a darkened room. Although Lombard speech occurs in the absence of a conversational partner, it is significantly modulated by communicative intent (Garnier et al., 2010). Since exploring communicative adaptations is of critical interest here, it is important to develop more

interactive experimental paradigms— perhaps allowing the participant to directly speak to a partner in the control room via audio or video link-up.

Overall, these findings demonstrate that masking sounds do not solely affect speech production mechanisms by reducing the talker's ability to self-monitor. Instead of the emphasis on self-monitoring seen in many studies of speech production (Christoffels et al., 2007; Lind, Hall, Breidegard, Balkenius, & Johansson, 2014), these results suggest that perceptual systems are also processing information in our acoustic surroundings, such that there is a route for meaningful elements in unattended auditory streams to be processed centrally. Indeed, auditory streams that are high in informational content (or semantic content) are processed centrally even when the task at hand requires that we actively disregard it. Although the STG may function as an error monitor in some circumstances, this study did not find strong evidence in support of that conclusion. Further studies with more sensitive analysis techniques may be able to establish whether we are seeing a role for multiple auditory streams of information in STG associated with both production and perception mechanisms, as has been previously suggested for perception (Rauschecker & Scott, 2009; Zatorre, Bouffard, Ahad, & Belin, 2002). Meanwhile, this study emphasizes the importance of not assuming that the STG is solely focused on error detection and audibility during speech production -- and not underestimating the effect that informational content has on us when we attempt to speak in background noise.

CHAPTER 6: STUTTERING AND SYNCHRONIZED SPEECH

6.1. ABSTRACT

This study tested the idea that stuttering is caused by over-reliance on auditory feedback. The theory is motivated by the observation that many fluency-inducing situations such as synchronised speech alter or obscure the talker's feedback. Typical speakers show 'speaking-induced suppression' in STG when speaking compared to listening. People who stutter may lack this response. In an fMRI scanner, people who stutter spoke in synchrony with an experimenter, in synchrony with a recording, on their own, in noise, listened to the experimenter speaking and read silently. No speech suppression response was observed. However, there was strong activity in STG in response to the synchronised speech condition, in which all participants spoke fluently. An independent component analysis confirmed that synchronised speech conditions significantly modulated activation in bilateral superior temporal gyri. Meanwhile, activation in a network of subcortical regions including the basal ganglia was modulated by speaking alone and speaking in masking noise, but not by synchronizing. The percentage of stuttered syllables was correlated with activation in frontal cortex and cerebellar vermis.

6.2. INTRODUCTION

Stuttering is a speech disorder characterized by frequent sound prolongations and syllable repetitions. Most children who stutter recover before puberty, but for 20% of those affected the disorder persists into adulthood, and while several speech therapy programmes and private courses offer techniques to manage speech and reduce the effects of disfluency, there is currently no 'cure' for the disorder: children who still stutter at the onset of puberty will likely stutter for the rest of their lives. There are, however, some manipulations that appear capable of temporarily inducing fluent speech in people who stutter- for example, speaking in synchrony with another person, or with pitch shifted auditory feedback. Finding out what links these fluency-enhancing conditions may provide insight into the disorder. The experiment described in this chapter was intended to test the theory that stuttering is caused by an overreliance on auditory feedback by comparing the neural and behavioural effects of two different types of altered auditory feedback— choral speech and masking noise. To put this experiment into context, an overview of the major behavioural and neural characteristics of stuttering is given below, followed by a description of the theory tested in this chapter and others that have sought to explain the disorder.

WHAT IS STUTTERING? WHO STUTTERS?

Because there is no genetic or medical test for stuttering, the disorder is defined behaviourally. One difficulty in classifying the speech of people who stutter, however, is that typical speakers are also often dysfluent. The DSM-IV characterizes stuttering in terms of "frequent repetitions or prolongations of sounds or syllables" (American

Psychiatric Association, 2000): these stuttering incidents can be further divided into stallings (where speakers pause or repeat part of the speech already uttered) and advancings (repeating the first syllable or phoneme of the next word). But of course, similar disfluencies are part of everyday conversational speech even for people with fluent speech. Wingate (1964) suggested a more comprehensive, three-part definition of the disorder, encompassing verbal expressions (repetitions, prolongation), 'accessory features' (that is, physical concomitants such as tics, involuntary hand movements) and psychological impact (e.g. anxiety). However, there is no diagnostic test currently available that integrates psychological and physical manifestations of stuttering. To evaluate stuttering severity in this experiment, we used the Stuttering Severity Instrument IV (Riley, 1972), which gives a severity score based on the severity of physical tics, and the number and duration of stuttering incidents. For the purposes of the SSI-IV, a stuttering incident is defined as a repetition or prolongation of a sound or syllable, including silent prolongations or 'blocks' (long pauses before a word). Whole word repetitions are not counted as stuttering incidents unless the word is monosyllabic, because word repetitions occur frequently in typical speech so may not be a pathological disfluency. This is the most reliable diagnostic test currently available: the scores have been standardized using a sample of 109 children and 28 adults, and the test has high inter-rater and test-retest reliability (Riley, 1972).

Although the aetiology of stuttering remains unclear, there is generally good epidemiological evidence about its prevalence. In a landmark longitudinal study of 1024 families, Andrews and Harris (1964) found that around five percent of children experienced developmental dysfluency, with the onset of the disorder usually occurring

early but not concurrently with language- the two most commonly reported ages at which children began to stutter were two and five years old, although stuttering could begin at any age up to around 11 years. At the end of the study period of 15 years, 80% of the affected children had recovered, whilst 20% (i.e., 1% of the whole cohort) had not. Work on adults has showed that approximately 1% of adults stutter, implying that those children whose dysfluency persists into teenage years do not recover later in life (Bloodstein, 2006; Dworzynski et al., 2007; Månsson, 2000; Yairi & Ambrose, 1999). Boys are more likely than girls to persist in stuttering, while females recover earlier (Yairi & Ambrose, 1999; Dworzynski et al., 2007): in childhood, the ratio of boys to girls is 2.4:1, increasing to 4:1 in adulthood (Andrews & Harris, 1964).

WHAT BEHAVIOURAL SITUATIONS IMPROVE FLUENCY?

While adults who stutter are not likely to permanently recover, they may experience periods of fluency. One of the most comprehensive surveys of fluency-inducing situations was carried out by Bloodstein (1950), who interviewed 50 people who stutter about circumstances that might affect their speech. Participants were given 115 different situations and asked to rate their speech behaviour in each situation from 1 to 4, with 1 being as much (or more) stuttering than usual, and 4 being no stuttering at all. The effectiveness of each situation in inducing fluency was evaluated based on the percentage of participants that rated their speech as either 'Hardly any stuttering, or very markedly less' (3 on the rating scale) or as 'No stuttering at all' (4). Participants reported improved fluency in a variety of situations, from 'speaking to an animal' to 'When under fire during the war'. Some examples of particularly highly rated situations were 'Reading in unison with others who are reading different material' (which markedly improved or eliminated

stuttering in 76.5% of respondents) and speaking in time with rhythmic activities such as a swing of their arm (92.3% of respondents), twisting their wrist (94.1%) or walking (82.4%). However, in general responses were highly variable, and just two out of 115 situations resulted in improved or totally fluent speech in 100% of respondents: singing, and 'reading aloud in unison with others who are reading the same material'— also known as choral or synchronous speech.

Many of the situations that Bloodstein (1950) identified as most reliably improving stuttering can be classified as affecting either auditory feedback, or speech rate. Several other studies have confirmed that synchronous speech— which affects both timing and auditory feedback— results in a highly consistent and often dramatic reduction in stuttering (Andrews, Howie, Dozsa, & Guitar, 1982; Barber, 1939; Kalinowski & Saltuklaroglu, 2003) and the resultant increase in fluency is usually greater than that demonstrated under other fluency-enhancing conditions (Johnson & Rosen, 1937; Kiefte & Armson, 2008). In subsequent research, masking noise, frequency altered feedback and delayed auditory feedback have also been associated with improved fluency in people who stutter. For example, in a study of 54 PWS, Cherry and Sayers (1956) found that while blocking participants' ears was insufficient to reliably ameliorate stuttering, masking talkers' voices with a loud tone of 140Hz resulted in an immediate increase in fluency. White noise also significantly reduces stuttering when presented binaurally (Brayton & Conture, 1978; Murray, 1969; Yairi, 1976) and the increase in fluency is proportional to the amount of speech that is masked (Burke, 1969). Sutton and Chase (1961) found that this fluency-enhancing effect occurred not just when continuous noise masking was present, but when the noise was only triggered while participants were

speaking and, remarkably, when noise occurred only in silent gaps in the participants' speech. However, because the white noise was triggered by the talker's voice, there was an overlap between the start of articulation and the cessation of white noise that means participants were not producing speech in total silence. A study aimed at eliminating this confound (Altrows & Bryden, 1977) was unable to replicate the fluency-inducing effect of white noise bursts preceding speech. Additionally, the effect of masking noise on fluency is not totally consistent, and in one case has even been reported to increase the incidence of stuttering when white noise was presented for longer than five minutes (Garber & Martin, 1974).

Delayed auditory feedback (DAF) and frequency altered feedback (FAF) both induce fluency more effectively than masking noise (Kalinowski, Armson, Stuart, & Gracco, 1993). Delayed auditory feedback has the unusual effect of inducing disfluencies in typical speakers- although it should be noted that this speech can be reliably distinguished from 'true' stuttering 92% of the time, according to Neelley (1961). The 'fluent' speech of PWS experiencing DAF can also be distinguished from the fluent speech of typical speakers, but to a lesser degree (66% of the time). DAF and FAF are equally effective at reducing disfluency (Kalinowski et al., 1993; Macleod, Kalinowski, Stuart, & Armson, 1995) but affect talkers differently: subjects raise their vocal intensity during DAF but lower it when experiencing FAF (Peter Howell, 1990), and prolong vowel duration under DAF but not FAF (Peter Howell, El-Yaniv, & Powell, 1987). This suggests that the increase in fluency is not a side-effect of the feedback manipulation caused by incidental changes to speech intensity or vowel duration, but is the result of some other common property of the two techniques. However, there remains considerable individual

variability in responses to both DAF and FAF: Natke (2000) found that while DAF resulted in a reduction in stuttering frequency and duration in twelve subjects, FAF had no significant effect on mean fluency, while Armson and Stuart (1998) found that FAF resulted in a significant decrease in stuttering frequency during reading but not during spontaneous speech, and Sparks et al. (Sparks, Grant, Millay, Walker-Batson, & Hynan, 2002) reported that DAF induced fluency in severely dysfluent subjects but not in mildly dysfluent participants.

Andrews et al. (1982) found that PWS's fluent speech under altered auditory feedback conditions was characterised by slowed speech rate or lengthened phonation duration, leading to the suggestion that fluent speech is effected by changes in speech timing and phonation rather than specifically by feedback alteration (Brayton & Conture, 1978; Costello Ingham, 1983) Decreased speech rate on its own is reliably associated with an increase in fluency (Adams, Lewis, & Besozzi, 1973; Bloodstein, 1948; Perkins, Kent, & Curlee, 1991) as is 'metronome speech', in which talkers match their utterances to an external pacing signal such as a metronome beat (Toyomura, Fujii, & Kuriki, 2011). While some research has found that PWS are less fluent under frequency altered feedback when they speak at a fast speech rate compared to their normal speaking rate (Hargrave et al., 1994), other studies suggest that both FAF and DAF improve fluency at both normal and fast speech rates (Kalinowski et al., 1993; Kalinowski, Stuart, Sark, & Armson, 1996; Macleod et al., 1995; Sparks et al., 2002). This suggests that the fluency enhancing effect of altered feedback cannot be completely attributed to decreased speaking rate. In this experiment, two altered auditory feedback conditions were investigated: masking noise, and synchronous speech. Contrasting masking noise (arguably the least reliably fluency-

enhancing altered feedback technique) with synchronous speech (one of the most reliable manipulations) allows us to look closely at differences between the two conditions that may explain differences in their efficacy, including the speech rate they induce.

How are the brains of people who stutter different to those of typical speakers?

Theories about why certain behavioural manipulations affect the speech of PWS can be supplemented with neuroimaging data comparing fluent and dysfluent speech, while comparing the brains of PWS to those of controls may shed light on the origins of the disorder. Because people who stutter are hard to recruit and neuroimaging studies are hard to conduct, neuroimaging studies of people who stutter typically have only a small number of participants and may be underpowered. For this reason, it is most helpful to look at meta-analyses that can aggregate large numbers of studies and find common areas of activation overlap. Brown et al. (Brown, Ingham, Ingham, Laird, & Fox, 2005) performed an ALE (Activation Likelihood Estimate) meta-analysis of functional neuroimaging studies that had looked at adult people who stuttered. They found that the findings fell into three groups. First, stutterers displayed overactivation of cortical motor areas such as primary motor cortex and the supplementary motor area relative to controls. Secondly, the brains of stutterers showed anomalous lateralization- bilateral or right-dominance of speech-related areas that are usually left-lateralised in typical speakers. Finally, stuttered speech resulted in suppression of auditory areas that are usually active during speech production. From these general observations, Brown distilled three specific 'neural signatures of stuttering': (1) cerebellar vermis over-activation; (2) Bilateral absence of auditory activation in STG and (3) Activity in the right

frontal operculum and anterior insula. One drawback of Brown et al's (2005) original analysis is that it did not distinguish between studies that compared PWS with controls (treating stuttering as a 'state') and those that compared PWS's fluent speech to their dysfluent speech. Two recent meta-analyses (Belyk, Kraft, & Brown, 2015; Budde, Barron, & Fox, 2014) have addressed this by replicating Brown's results and then breaking them down into 'trait' and 'state' characteristics of stuttering. In general, findings fall into three categories, which approximately align with Brown's three neural signatures: subcortical structures; auditory and motor cortex; and anomalous hemispheric lateralization. The results of functional and structural analyses are discussed below with reference to these three broad categories.

SUBCORTICAL STRUCTURES: BASAL GANGLIA AND CEREBELLAR VERMIS

Although subsequent meta-analyses have confirmed Brown's assertion that overactivation of the cerebellar vermis is characteristic of stuttering, accounts differ as to whether this overactivation is associated with 'trait' or 'state' stuttering. In their meta-analysis, Budde et al (2014) found that dysfluent speech was associated with increased activation in left cerebellar vermis compared to fluent speech. Belyk et al. (2015) by contrast found that overactivation of the cerebellar vermis was associated with state stuttering (i.e. when PWS were compared to controls) but found less activation in the cerebellar vermis during dysfluent speech when compared to fluent speech. The vermis is mainly involved with postural adujustments such as the rhythmic modulation of walking movements, and is connected to spinal motor neurons via the vestibular nuclei and the reticular formation; it is possible that overactivation may represent a problem in movement timing. Along similar lines, other studies have suggested that stuttering may

be caused by abnormalities in the basal ganglia: one study (Lu et al., 2010) found significant differences in grey matter volume in the basal ganglia-thalomocortical circuit, while Alm and Risberg (2007) reported that a high proportion of subjects who stuttered had past incidents involving head injury, arguing that this could have caused damage to the basal ganglia. This is supported by the fact that basal ganglia disorders such as Parkinson's disease often lead to the re-occurrence of developmental stuttering, even when the patient had previously recovered (Shahed & Jankovic, 2001) Giraud et al. (2008) found that stuttering severity was negatively correlated with activity in the basal ganglia. Studies have found attenuated structural and functional connectivity in the basal ganglia-thalamocortical loop in both children (Chang & Zhu, 2013) and adults (Lu et al., 2010), with specific weakness in connectivity between the left posterior middle temporal gyrus and putamen (Lu et al., 2010), while PWS had stronger connections from thalamus to putamen and pre-SMA; however, children exhibited less connectivity between putamen and SMA.

AUDITORY AND MOTOR CORTEX

It seems obvious that trait stuttering (i.e stuttering compared with fluency) should result in greater activity in motor cortex- after all, PWS are moving their mouths more when they stutter than when they do not. However, there is some evidence of abnormalities in the structure and function of motor cortex in PWS that may plausibly cause stuttered speech, rather than result from it. For example, Chang et al. (Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008) found that children with persistent developmental stuttering had reduced white matter connectivity in tracts underlying motor areas for the face and larynx. Additionally, stuttering has been associated with unusual activation in

the supplementary motor area (which forms part of a circuit regulating motor timing with the basal ganglia). Both Budde et al. (2014) and Belyk et al (2015) found that studies' activation foci significantly converged in the SMA, although the two meta-analyses are in disagreement about whether this convergence represents trait stuttering (Budde et al, 2014) or state stuttering (Belyk et al, 2015). However, Budde et al's (2014) analysis revealed that state stuttering was associated with activation in the pre-SMA, which forms part of the same cytoarchitectonic area as the SMA, BA6- although there are several differences in the function and structure of the two regions- for example, SMA is associated with movement generation and control, while pre-SMA is involved in higher-order processes such as action preparation (Lima, Krishnan, & Scott, 2016). Small differences in results are most likely explained by the fact that each meta-analysis looked at a different, though overlapping, set of studies— six studies analysed by Belyk et al. were not included in Budde et al's meta-analysis, while four studies in Budde et al's analysis were omitted by Belyk et al.

People who stutter also demonstrate reduced left-lateralization of motor activation compared to controls. For example, during transitions between speech gestures, the excitability of left tongue motor cortex is enhanced in typical speakers but not in people who stutter (Neef, Hoang, Neef, Paulus, & Sommer, 2015). Additionally, during single word reading, fluent talkers activate left inferior frontal cortex prior to left motor cortex, but stuttering participants show the opposite sequence of activation, suggesting an impairment in communication between left motor cortex and IFG.

Further evidence of impaired projections from inferior frontal cortex in PWS comes from connectivity research showing that the arcuate fasciculus, which links auditory cortex to

the motor system via the inferior frontal gyrus, is degraded in people who stutter (Chang et al., 2008; Connally, Ward, Howell, & Watkins, 2014) A meta-analysis of diffusion tensor imaging studies (Nicole E. Neef, Anwander, & Friederici, 2015) found that PWS display reduced white matter integrity in the left superior longitudinal fasciculus (including part of the arcuate fasciculus), specifically in tracts connecting inferior frontal cortex with the parietal cortex (including angular gyrus) and the superior and middle temporal gyri in temporal cortex. Foundas et al. (2004) found that stutterers whose speech became more fluent when speaking under delayed auditory feedback also had atypical planum temporale asymmetries, while stutterers who did not have atypical planum temporale asymmetries also did not become more fluent when speaking under delayed auditory feedback. This area of the brain has been implicated in auditory-motor integration (Hickok, Buchsbaum, Humphries, & Muftuler, 2003), suggesting that in some stutterers the condition may be associated with disordered auditory-motor interfaces. Additionally, many sources of evidence suggest that the auditory cortex itself is underactive in people who stutter (Wu et al., 1995). Decreased bilateral STG activation relative to controls was, of course, one of Brown's 'neural signatures' of stuttering. Budde et al. (2014) and Belyk et al. (2015) both found that the right STG was less active in people who stutter compared to controls ('trait' activation), while the left STG was underactive during stuttering compared to fluent speech ('state' activation). The bilateral absence of activation reported by Brown et al. (2005) therefore indicates a dysfluent state in someone with a stuttering trait.

ANOMALOUS RIGHT-LATERALIZATION AND BRAIN ASYMMETRIES

In addition to the unusual suppression of activity in STG during speaking discussed above, another study found anomalous lateralization of activation in the auditory cortex during listening in PWS (Sato et al., 2011). This study used near-infrared spectroscopy (NIRS) to look at regional changes in blood flow while adults and children who stutter listened to Japanese single word stimuli pairs that differed either in prosody (/itta/ versus /itta?/) or in phonology (/itta/ versus /itte/). While controls showed a left-dominated response to phonemic contrast and a right dominated response to prosodic contrasts, most of the stuttering subjects displayed no difference in lateralization between conditions, and those that did showed the opposite pattern of lateralization to controls. The subjects in this study included children as young as three, meaning that the atypical brain responses seen here are likely to be a true symptom or cause of stuttering, rather than a reflection of coping mechanisms.

Structurally, several abnormalities have been reported in the brains of people who stutter- for example, extra sulci in the pars opercularis in the inferior frontal gyrus (Foundas, Bollich, Corey, Hurley, & Heilman, 2001) and in the second segment of the right lateral fissure (Cykowski et al., 2008). One common finding is that the brains of people who stutter have reduced left hemisphere grey matter compared to controls, for example in the rolandic operculum (Sommer, Koch, Paulus, Weiller, & Büchel, 2002) and the inferior frontal gyrus (Chang et al., 2008; Kell et al., 2009). People who stutter may also lack brain asymmetries found in people who do not stutter: some (Foundas et al., 2001) but not all (Foundas et al., 2004, Cykowski et al. 2007) people who stutter have reduced asymmetry in the right and left planum temporale compared to controls. Moreover, some

people who stutter do not have asymmetries in the size of prefrontal cortex (typically larger in the right than left hemisphere) and the occipital lobe (typically larger in the left than the right hemisphere) (Foundas et al., 2003). However, results are inconsistent, with one study reporting no difference in grey matter volume between people who stutter and controls (Jäncke, Hänggi, & Steinmetz, 2004)and another finding no unusual hemispheric asymmetries in the brains of children who stutter (Chang et al., 2008), suggesting that structural abnormalities may arise as a result of compensation for the stutter, rather than as a result of stuttering itself.

THE NEURAL CORRELATES OF INDUCED FLUENCY

Based on the evidence shown above, we can draw some general conclusions about the areas of the brain that are involved in fluent and dysfluent speech. When the dysfluent speech of people who stutter is compared to their speech in fluency-enhancing conditions, disfluency is associated with activation in bilateral inferior frontal gyri, motor and somatosensory cortex, basal ganglia and cerebellum. Fluent speech, on the other hand, is associated with activation in the right superior temporal gyrus and bilateral middle temporal gyrus. (Budde et al, 2014; Belyk et al., 2014). That is, fluency enhancement is associated with decreased activity in motor areas that are normally over-active, and increased activity in auditory areas that are usually under-active (P. T. Fox et al., 1996). For example, Watkins et al. (Watkins, Smith, Davis, & Howell, 2007) found that during normal feedback (compared to controls), PWS had lower activity in left ventral premotor cortex, right central opercular cortex, left and right sensorimotor cortex and left anteromedial Heschl's gyrus. In both groups, the two types of altered auditory feedback were associated with increased activation in bilateral superior temporal cortex

in comparison to normal feedback, while delayed feedback resulted in an increase in right inferior frontal cortex activation when contrasted with normal feedback. However, there was no significant difference between the responses to altered auditory feedback between PWS and controls, suggesting that as PWS became more fluent, their brain activity converged on that of controls.

WHAT DO RESEARCHERS BELIEVE CAUSES STUTTERING?

We saw that stuttering is associated with structural and functional abnormalities in the auditory and motor cortex, in subcortical areas involved in timing of movement, and in hemispheric lateralization. Based on this evidence, researchers have proposed variously that stuttering arises from a deficit in feedback processing, a problem with motor timing, or inefficient interhemispheric communication. Below, each of these ideas is evaluated with reference to relevant research.

One of the earliest accounts of stuttering attributed the disorder to atypical lateralization in the brain (Geschwind & Galaburda, 1985; Travis, 1978). This hypothesis, called the cerebral dominance theory, suggests that stuttering could be caused by the left hemisphere having weak or incomplete dominance over language, resulting in 'confused laterality'. According to this theory, the left and right hemisphere struggle to gain dominance of speech processing, leading to less efficient speech processing. As handedness reflects language dominance (people who are right-handed generally have left-hemisphere dominance for language, and vice versa), it was argued that stuttering could be caused by switching handedness- for example, by forcing a left-handed child to write with her right hand (Bryngelson, 1935; Milisen & Johnson, 1936). However,

stuttering emerges before children learn to write (Proctor et al., 2008), and later research has not supported the idea that stuttering is linked to handedness (Rosenfield, 1980). In addition, the idea that stuttering is caused by atypical interhemispheric relationships has been contradicted by recent research comparing interhemispheric inhibition in people who stutter with controls (Sommer et al., 2009). Interhemispheric inhibition (IHI) describes the interplay between left and right motor cortices that is necessary to produce unilateral movement such as writing with one hand without echoing its movements with the other. When motor cortex on one side of the brain is stimulated, it sends an inhibitory signal to its partner on the other side. Sommer et al (2009) found no difference in IHI between PWS and controls, suggesting that communication between hemispheres is intact.

However, it remains possible that atypical lateralization may play a role in causing stuttering. For example, if speech is primarily controlled by the right hemisphere, this may be inadequate for processing language. Alternatively, it is possible that speech begins normally in the left hemisphere but is diverted through the right hemisphere, leading to inefficient processing. Research on adult stutterers who have benefited from stuttering therapy has found that these adults show increased left hemisphere activity when speaking (Neumann et al., 2005). Thus, it is possible that recovery occurs when the stutterer's brain is able to reorganize speech pathways, while for some reason this reorganization never happens in the brains of persistent stutterers. However, other research has shown that activity in the left hemisphere auditory cortex is correlated with dysfluent speech while right hemisphere auditory cortex overactivity is associated with fluent speech (Braun et al., 1997; Neumann et al., 2003); this result, which has also been

confirmed by the meta-analyses discussed above, suggests that right-hemisphere activation is actually beneficial or compensatory. This is consistent with research showing that aphasic patients are use right hemisphere homologues of left hemisphere language regions after experiencing infarcts affecting speech production in the left hemisphere, although this is not necessarily associated with successful compensation for the language deficit (Price & Crinion, 2005). With this evidence in mind, it seems most likely that atypical lateralization arises from compensation for stuttering, rather than causing the disorder itself.

An alternative group of theories has formed based on the observation that many stuttering therapies work by altering the auditory feedback signal. People who stutter report improvements when speaking in delayed auditory feedback, frequency shifted feedback, speaking in noise, and speaking in chorus with others. However, when the masker or altered feedback stops, stuttering resumes. This, in conjunction with the functional anomalies seen in auditory cortex in neuroimaging studies of people who stutter, has led some researchers to suggest that stuttering is a manifestation of a central auditory processing disorder (Salmelin et al., 1998), or some difficulty with speech monitoring. Max et al. (Max, Guenther, Gracco, Ghosh, & Wallace, 2004) suggested two possible hypotheses, based on Guenther's (2006) DIVA model of speech production. First, the underlying internal speech models in people who stutter may be poorly specified in some way. Under this hypothesis, stuttering manifests itself during development when children are unable to update their internal speech models appropriately in response to feedback, and may have a problem with accessing or forming mappings between motor commands and sensory responses. As a result their internal model is mis-specified and

sends inaccurate feedforward commands to the articulators. The mismatch between the faulty prediction and the actual sensory consequences of the executed movement results in attempts to correct the speech by reissuing the motor command, resulting in stuttering. Altered feedback induces stuttering, therefore, because it activates auditory cortex and stimulates the internal model. Another model that develops the idea of stuttering as an internal model deficit is the Covert Repair Hypothesis, or CRH (Postma & Kolk, 1993). The CRH uses Levelt's (1983) three-loop monitoring system as a theoretical frame, rather than Guenther's DIVA model; however, Levelt's internal monitoring loop (defined as the inspection of the articulatory plan) and Guenther's feedforward loop are both conceptually similar in that they describe a stage of speech monitoring that occurs before articulation. The Covert Repair Hypothesis suggests that disfluencies arise because the speaker has detected an error during internal monitoring and is attempting to correct it. Such 'covert repairs' occur more frequently in people who stutter owing to a deficit in phonological encoding. This theory is based on the spreading-activation account of phonemic control (Dell, 1986), in which word selection is accomplished by activating all phonemic, semantic and syntactic nodes associated with the word; this activation spreads to surrounding nodes until the most highly activated node is selected. If selection occurs too early, it is more likely that the wrong node will be selected, leading to an error, which in turn triggers a covert repair. Postma & Kolk (1993) argue that PWS are slow to activate the right representation, so are more likely to make these errors. However, evidence does not support the idea that PWS have a phonological disorder: children who stutter make the same amount of phonological errors as fluent children, and the number of phonological errors made does not correlate with stuttering severity

(Nippold, 2002). Additionally, evidence suggests that adults who stutter do not have a slower rate of phonological encoding compared to fluent adults (Brocklehurst, 2008).

As an alternative, the second theory put forward by Max et al. (2004) involves no problems with the internal model. Rather, PWS may have weakened feedforward projections and are thus forced to rely on feedback monitoring. Overreliance on feedback monitoring results in system resets and effector oscillations as the talker attempts to compensate for the time delay between the motor command being issued and the feedback being received. Under this hypothesis, altered auditory feedback prevents the talker from relying on the feedback circuit and encourages them to use the weakened feedforward projections, resulting in fewer corrections. A computational modelling study testing this theory using the DIVA model (Civier, Tasko, & Guenther, 2010) found that programming the model to rely more on auditory feedback resulted in more acoustic errors than the default model parameters, particularly during rapid formant transitions. The model did not produce stuttered speech with these parameters; however, the auditory errors produced are consistent with those found in human studies (Blomgren, Robb, & Chen, 1998; Robb & Blomgren, 1997) which found that people who stutter have significantly lower F2 values when producing syllables with rapid formant transitions compared to syllables without rapid transitions. The authors (Civier et al, 2010) suggest that an accumulation of these errors eventually causes a system reset, in which the syllable is restarted, leading to sound and syllable repetition. One interesting observation that may support the theory that stuttering is related to over-reliance on auditory feedback is the existence of several surveys suggesting that there is a much lower

incidence of stuttering in the deaf population than in the population at large, although this evidence is largely anecdotal (Backus, 1938; Harms & Malone, 1939; Wingate, 1970).

The EXPLAN model (Howell & Au-Yeung, 2002) takes a different approach by suggesting that planning (linguistic processing) and execution (motor processing) are two independent systems that control the production of spontaneous speech. When linguistic plans are sent to the motor system too late, dysfluency results. Stutters are therefore seen as attempts to allow the motor system to catch up. Thus, 'stallings'- repeating part of the last utterance- are interpreted as 'playing for time' while the motor system waits for the next bit of the linguistic plan to be delivered. Meanwhile, 'advancings'—stutters on the first syllable or phoneme of words—result from an attempt to go ahead with the next word in the hope that the rest of the plan will be delivered by the time the first syllable has been executed. When typical speakers are required to provide a fast commentary on a cartoon, the number of part-word repetitions in their speech increases (Howell & Sackin, 2000) in keeping with EXPLAN's prediction that stuttering occurs whenever the articulation rate exceeds the speech planning rate. However, it is arguable whether the experiment induced 'true' stuttering in fluent speakers, especially as subjects did not produce silent blocks, which are a major feature of most stuttered speech.

This account could also explain why both timing regulation techniques and altered feedback sometimes induce fluency, since one common effect of all such techniques is that they slow speech rate, which theoretically could allow enough time for linguistic plans to be fully delivered to the motor system in time for the next utterance. However, as previously mentioned, there is evidence that altered feedback techniques ameliorate stuttering independent of their effect on speech rate, suggesting that accounts which rely

on speech rate alone are unlikely to fully explain stuttering. In general, attempts to link stuttering to a timing deficit have been largely inconclusive: for example, Max and Yudman (2003) asked participants to synchronize movements to a regular auditory stimulus and then continue the pattern when the stimulus ended, but found no significant difference between PWS and controls for syllable vocalization, nonspeech lip movement, or finger movement.

STUDY JUSTIFICATION

SYNCHRONOUS SPEECH PRODUCES RELIABLE ADAPTATION IN TYPICAL AND ATYPICAL SPEAKERS.

The experiment described in the following chapter aimed to test the theory that stuttering is caused by an over-reliance on auditory feedback by using fMRI to measure blood flow while people who stutter spoke in quiet and with different kinds of altered feedback- specifically, choral speech and masking noise. Like masking noise, synchronous speech occurs in many real-life contexts among typical speakers. For example, synchronous speech is frequently an integral part of activities that promote social cohesion, such as praying or reciting oaths of allegiance. People are able to synchronise with each other rapidly and without rehearsal (Cummins, 2003) even when the text or message they are repeating has no obvious metrical structure (such as that found in nursery rhymes) (Cummins, 2009; King, 2012).

The study is closely based on an fMRI study by Jasmin et al. (2016), which looked at the effects of synchronous speech in typical speakers. Subjects read sentences in synchrony with an experimenter, on their own, and while the experimenter read a different sentence (asynchronous speech); they also listened to sentences being read and passively

observed a fixation cross. In a covert manipulation, in half of the synchrony trials the experimenter's voice was pre-recorded rather than live. Results showed that neural and behavioural responses to synchronising with a live speaker could be reliably dissociated from responses to synchronising with a recording, even when participants were not aware that they were speaking with a recording. People are able to synchronise more closely when they are speaking with a live speaker than with a recording. Neurally, activity in the right temporal pole was significantly attenuated when speaking alone, synchronising with a recorded speaker and speaking over a recording of a different text compared to listening. However, there was increased activation in this region when talkers synchronised with a live speaker. These results were interpreted as indicating a release from speaking-induced suppression when reciprocally synchronising with a partner, treating the participant's voice as equivalent to an external stimulus. It was hypothesised that, by blurring the boundary between self- and other-produced stimuli, this response could reflect the feeling of social cohesion promoted by participating in a synchronised activity. In this study, although the primary aim was to investigate the relationship between type of feedback, neural activation and fluency, we were also interested in whether the same distinction between synchronising with a recording and synchronising with a live partner could be found in people who stutter.

COMPARING FEEDBACK TYPES CAN SHED LIGHT ON STUTTERING

We included a white noise masking condition in place of Jasmin et al's (2016) asynchronous speech condition. Although other studies have compared different types of auditory feedback (e.g. Watkins et al, 2007 compared delayed and frequency-shifted feedback), these were techniques that are approximately equivalent in their effectiveness

at inducing fluent speech. Masking noise induces fluency in some people who stutter, but does not do so as reliably as synchronous speech. By contrasting these two techniques, we aim to investigate whether there are neural or behavioural characteristics of participants' responses to the two feedback types that could explain differences in their effectiveness. For example, if fluency is related to speed of articulation, then we would expect to find that participants' speech rates are consistently slowed by synchronous speech, but not by masked speech. Alternatively, if stuttering results from an overreliance on auditory feedback, we might expect differences in superior temporal gyrus activation between the two conditions.

Testing the feedback overreliance theory of stuttering

PWS frequently demonstrate increased right hemisphere activation in regions of the precentral and inferior frontal gyrus associated with some responses to perturbed feedback in typical speakers; this has been proposed as neural evidence supporting the idea that stuttering is caused by overreliance on auditory feedback (Tourville et al, 2008). It might, therefore, be expected that dysfluent speech is associated with increased activation in the superior temporal gyrus, which in typical speakers is associated with the processing of auditory feedback during speech production. However, previous reseach has established that stuttering is associated with underactivation in bilateral STG. Additionally, induced fluency has been shown to correlate with increased activity in the STG. It is possible that this reflects abnormalities in the auditory feedback loop. Alternatively, Jasmin et al's (2016) study suggests another way of interpreting this data. If suppression of activity in temporal cortex is a way of dissociating your own voice from that of others, then fluency may be associated with feelings of agency.

This study aimed to explore the implications of abnormal STG activation in stuttering further by looking at the effect of different types of auditory feedback on superior temporal cortex activation. We aimed to replicate the finding that fluent speech is associated with increased STG activation, and were interested in whether people who stutter display the speech suppression response characteristic of speech monitoring. Additionally, this study integrates neural and behavioural responses by recording participants' voices in the scanner and using fluency data (measured in percentage of syllables stuttered) as a covariate in the fMRI model.

6.3. METHODS

PARTICIPANTS

Participants were recruited through the British Stuttering Association and were adults who self-identified as people who stutter. 14 participants (5 female; mean age 38.7, s.d. 12.2, range 24-63) underwent behavioural pretesting to classify their stuttering severity and evaluate the effect of choral speech on their stutter. Participants were additionally screened for hearing loss using an Amplivox 116 Screening Audiometer with DD45 earphones (amplivox.ltd.uk). None of the participants met the critera for hearing loss (defined here as a four-frequency pure tone average threshold of more than 20dB).

They were invited back to participate in the fMRI study if they had a stutter of any severity as defined by the SSI-IV, and they became more fluent under choral speech conditions. One participant was excluded at this stage because they did not stutter during the behavioural test. Of those who were invited back, nine native British English speakers continued to the fMRI testing (1 female, mean age 34.7, s.d. 8.4, range 24-48).

Assessment for stuttering severity

Participants' speech was evaluated using the Stuttering Severity Instrument IV (Riley, 1972). The SSI-IV calculates a severity score based on the percentage of syllables stuttered in two speech tasks, the duration of the three longest stuttering incidents, and physical tics observed at the time of testing. Participants sat in a soundproofed room with two experimenters. One experimenter delivered the test materials while the second recorded information on physical concomitants. Participants' speech was recorded using a RODE NT1-A one-inch cardoid condenser microphone connected to a Windows computer via a Fireface UC high-speed USB audio interface (RME Audio, Haimhausen) Their voices were recorded at 44100Hz with 16 bit quantisation using Adobe Audacity 3.0.

Subjects spoke spontaneously for three minutes and read one of two passages aloud. The passages were either 369 or 374 syllables long. The other passage was used in the synchronous speech task and the order of the passages was counterbalanced across participants.

Synchronous speech outside the scanner

To evaluate the effects of synchronous speech on testing, participants read the second passage in unison with an experimenter positioned outside the testing room. The experimenter spoke into an AKG 190E cardoid dynamic microphone and heard through AKG K240 Studio on-ear headphones. This mimicked the effect of speaking in the scanner environment as the participant was unable to see their conversational partner and use

nonverbal cues. It additionally enabled us to record the participant's voice on its own, without the experimenter.

FMRI STIMULI

The fMRI paradigm was closely based on Jasmin et al (2016), with some difference in the technical setup and a speech in noise condition substituted for the 'Diff-Live' condition.

Participants lay supine in the scanner and saw sentences in yellow or blue on a black background projected onto an in-bore screen, using a video projector (Eiki International). They spoke into an OptoAcoustics FOMRI-III noise-cancelling optical microphone and heard stimuli through Sensimetrics S14 fMRI-compatible insert earphones. In the control room, the experimenter was seated in front of a RODE NT1-A 1" cardoid condenser microphone and heard the participant through Beyerdynamic DT100 circumaural headphones. The participant's voice, experimenter's voice and sound from the computer were routed through an RME Fireface UC 36-Channel, 24 Bit / 192 kHz USB high speed audio interface using TotalMix software and were recorded in three separate channels on a Mac computer. Routing was instantaneous, so there was no delay between the experimenter or participant speaking and their conversational partner hearing them.

FMRI TASK

The following five sentences were used as stimuli:

1. When sunlight strikes raindrops in the air, they act as a prism and form a rainbow.

2. There is, according to legend, a boiling pot of gold at one end of a rainbow.

3. Some have accepted the rainbow as a miracle without physical explanation.

4. Aristotle thought that the rainbow was a reflection of the sun's rays by the rain.

5. Throughout the centuries, people have explained the rainbow in various ways.

These sentences are adapted from The Rainbow Passage (Fairbanks, 1960), and were used as they are about the same length (mean syllables = 20.8 ± 1.3), and can be spoken comfortably during a short presentation window by typical speakers. It was expected that some participants who stuttered would not be able to complete the entire sentence in the six seconds allotted for the task, and this was factored into the analysis.

In every trial, participants saw a prompt telling them which condition was coming up next, followed by the text of one of the five sentences. Instruction prompts were displayed for three seconds, then replaced with a fixation cross which remained on screen for one second before the stimulus sentence was shown. There were six conditions:

1. Synch-Live: Participants saw a 'SYNCHRONIZE' prompt and read the sentence synchronously with the experimenter.

2. Synch-Rec: Participants saw a 'SYNCHRONIZE' prompt and read the sentence synchronously with a recording of the experimenter.

3. Speak-Noise: Participants saw a 'SPEAK IN NOISE' prompt and read the sentence over 83dB white noise.

4. Speak-Alone: Participants saw the prompt, 'SPEAK' and read the sentence on their own

5. Listen: Participants saw the prompt, 'LISTEN' and read the sentence silently while hearing a recording of the experimenter reading it aloud.

6. Read-Silently: Participants saw the prompt 'READ SILENTLY' and read the sentence silently with no auditory stimulus.

In the synchronization conditions, participants spoke with a male American English speaker, either live, through the microphone (Synch-Live) or recorded, via a laptop (Synch-Rec). Recorded trials in both Synch-Rec and Listen conditions were produced by the live experimenter during synchronous speech with a different partner. This was intended to isolate neural and behavioural correlates of speech with a live partner who can adaptively alter their voice to match yours (reciprocal synchronization) while controlling for auditory and motor requirements as closely as possible. The prompt for both synchronization conditions was identical apart from a colour code intended to tell the experimenter when live speech was required: the prompt text was yellow in the Synch-Live condition, and blue in the Synch-Rec condition. To disguise the colour code from participants, the colour of the prompts was varied randomly in all other conditions, so that the prompt was blue in half of all trials, and yellow in the rest. Post- test debriefing confirmed that none of the participants identified that they were synchronizing with a recording.

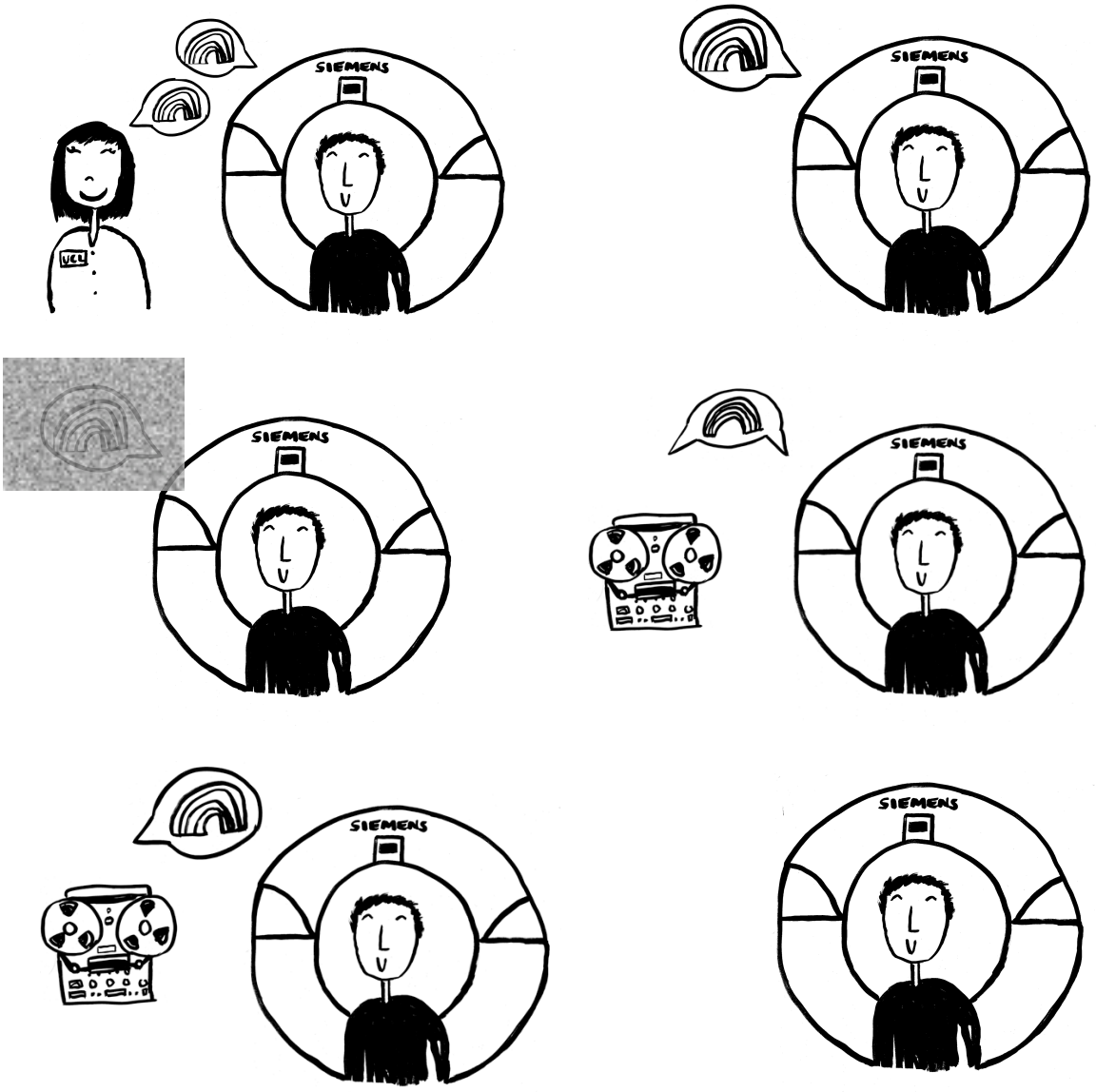


FIGURE 8: EXPERIMENTAL CONDITIONS (FROM TOP LEFT CLOCKWISE: SYNCLIVE, SPEAKALONE, SYNCREC, READSILENTLY, LISTEN, SPEAKALONE)

FMRI ACQUISITION PARAMETERS

Functional MRI images were acquired using a Siemens Avanto 1.5 Tesla scanner with 32-channel head coil, using a T2- weighted gradient-echo planar imaging sequence, which covered the whole brain (TR=9s, TA=3s, flip angle 90 degrees, 35 axial slices, matrix size=64x64x35, 3x3x3mm in-plane resolution). High-resolution anatomical volume images (HIRes MP-RAGE, 160 sagittal slices, matrix size: 224x256x160, voxel size=1 mm^3) were also acquired for each subject. Participants took part in three functional runs, each consisting of 55 trials (ten of each main condition and five ReadSilently trials). The five stimulus sentences were crossed with each of the six conditions, and combination of stimulus sentence and condition appeared twice per run. The order of the trial types was pseudo-randomized such that every five trials included one of each of the five stimulus sentences and one trial in each condition.

6.4. ANALYSIS

ACOUSTIC AND BEHAVIOURAL ANALYSIS

Two participants' recordings could not be used owing to problems with the recording setup. For the remaining seven participants, recordings of each experiment were divided up into individual trials using a MATLAB script. Each trial was evaluated for the total number of syllables, and the number of stuttering events, by a rater who was blind to the conditions. These scores were used to generate average speech rates (in syllables per second) and percentage of stuttered syllables for each participant in each condition.

The degree of synchrony in Synch-Live and Synch-Rec conditions was computed using a Dynamic Time Warping algorithm (Cummins, 2009). This converted the recordings to sequences of mel frequency-scaled ceptral coefficients, which were used to create a similarity matrix. The algorithm calculated how one speaker's utterance was warped in time relative to the other using a least-cost warp path through the matrix. A diagonal path would mean perfect synchrony; the larger the area under the warp path relative to the diagonal, the greater the degree of asynchrony on the trial.

Functional analysis

Preprocessing

First-level and group-level analysis was carried out using SPM 8. To allow for T1 saturation effects, the first three functional volumes of each run were discarded. Each participant's fMRI time series was realigned to the first volume of the run using six-parameter rigid-body spatial transformation and their mean functional image was coregistered to their anatomical T1 image; the scans were then re-oriented into standard space by manually aligning to the anterior commissure. The estimated parameters resulting from motion correction were inspected and did not exceed 3mm or 3 degrees in any direction. The T1 image was segmented into grey matter, white matter and cerebrospinal fluid; the parameters generated by this were used to spatially normalize the functional images into MNI space at $2mm^3$ isotropic voxels. The data was then smoothed using a Gaussian kernel of 8 mm^3 FWHM.

UNIVARIATE FUNCTIONAL ANALYSIS

At the single-subject level, events were modelled from the presentation of the stimulus sentence, using a canonical haemodynamic response function, with ReadSilently as an implicit baseline and motion parameters included as a regressor of no interest. Event duration was set at six seconds. Contrast images were calculated for each of the conditions using ReadSilently as a baseline, and for Synch-Live>Synch-Rec.

These contrasts were taken up to the group level and used to perform 1) a one-sample t-test for SynchLive>Synch-Rec; 2) a one-way repeated measures ANOVA looking at differences between each of the three speaking tasks (SpeakAlone, Synchronize, and SpeakNoise) compared to listening. Next, a multiple regression analysis was carried out on the subset of subjects for whom audio data was available (7 subjects) using the percentage of stuttered syllables in each trial as a regressor; this analysis revealed voxels that were more active when participant stuttered, regardless of the trial type. All contrasts were thresholded using a voxel wise familywise error rate correction for multiple comparisons at $p < 0.05$. Statistical images were rendered on the normalized mean functional image for the group of participants.

INDEPENDENT COMPONENT ANALYSIS

Spatial independent component analysis (sICA) was carried out using GIFT (mialab.mrn.org). Each subject's functional data was reduced in size using two steps of standard principal component analysis (PCA). The optimal number of components was estimated as 17 using the minimum description length critera. The Infomax algorithm was used to extract these 17 independent components and generate spatially independent BOLD maps and a time course for each one. This step was repeated 50 times

using ICASSO with a different random initiation seed each time, in order to assess the stability of the independent components. Next, individual subject spatial maps and time courses were back-reconstructed using information from the ICA and the data reduction stage, and used to generate statistical maps of group results.

Artefactual components were identified using a systematic procedure that combined spatial sorting and visual assessment using the process outlined by Griffanti et al. (2016). First, the average sICA maps were correlated with the probabilistic grey matter, white matter and cerebrospinal fluid (CSF) maps used for segmentation in SPM8. The spatial maps, power spectra and time series for each component were visually inspected and evaluated using the criteria given in Griffanti et al (2016). On the basis of the regression results and the visual inspection, the decision was made to exclude components if they correlated at $r>0.05$ with white matter, $r>0.19$ with CSF, or $r<0.01$ with grey matter. Nine independent components remained after the identification of artefacts. The maps were averaged across runs in each participant and used to conduct a random effects one-sample t-test in SPM8, thresholded at FDR $p<0.05$; this revealed which brain regions contributed to each component.

Components were temporally sorted using the SPM design matrices for each subject, which contained information about the onsets and time courses of each trial type. A multiple regression correlating the IC time courses and the modelled haemodynamic response function was carried out using the GIFT temporal sorting function. For each component, this generated beta-weights of each condition's correlation with the component time course (indicating increases and decreases in task-related activity in that component). These beta weights were averaged across sessions and subjects for each

condition, and statistically tested using SPSS (IBM) to assess differences between conditions.

6.5. RESULTS

BEHAVIOURAL PRETESTING

Participants were classified according to the SSI-IV using the recordings made during behavioural pre-testing, and represented a broad spectrum of stuttering severity from very mild to very severe (Fig 19).

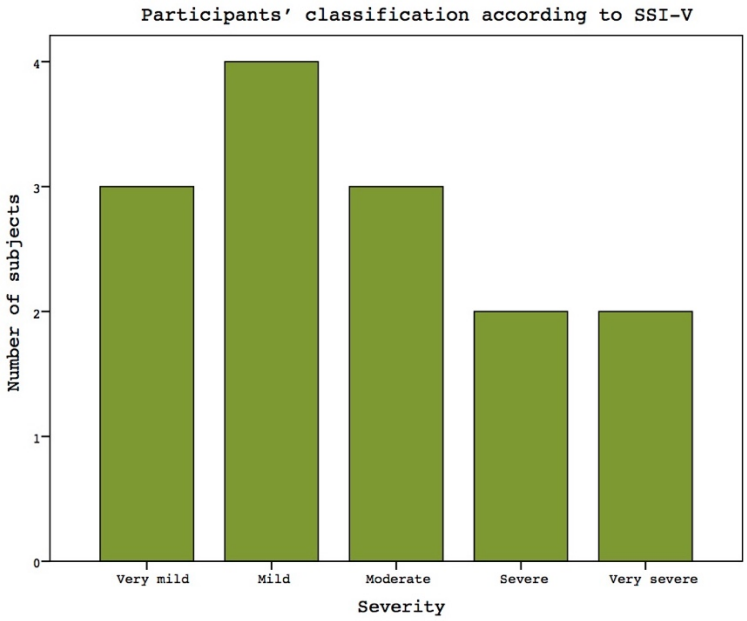


FIGURE 9: STUTTERING SEVERITY AS EVALUATED BY RILEY'S STUTTERING SEVERITY INSTRUMENT

A series of one-tailed t-tests were conducted to investigate differences in stuttering duration, frequency and speech naturalness. These revealed that when synchronizing, participants stuttered less than when speaking alone ($t(13)=2.149$, $p=0.025$, $d=0.51$), and the duration of the longest stuttering incident was significantly shorter when

synchronising (one-tailed t(13)= 1.987, p=0.034, d=0.48). However, there were no significant changes in speech naturalness between conditions (t(13)=0.099, p=0.46, d=0.02). In the speak-alone condition there was considerable variability in the percentage of stuttered syllables (mean= 7.5, s.d.=11.39) and duration of stuttering incidents (mean= 4.87, s.d.=7.48). By contrast, there was relatively little variability in participants' performance during choral speech, either in percent stuttered (mean=0.93, s.d.=0.73) or in duration (mean=0.83, s.d.=0.53)

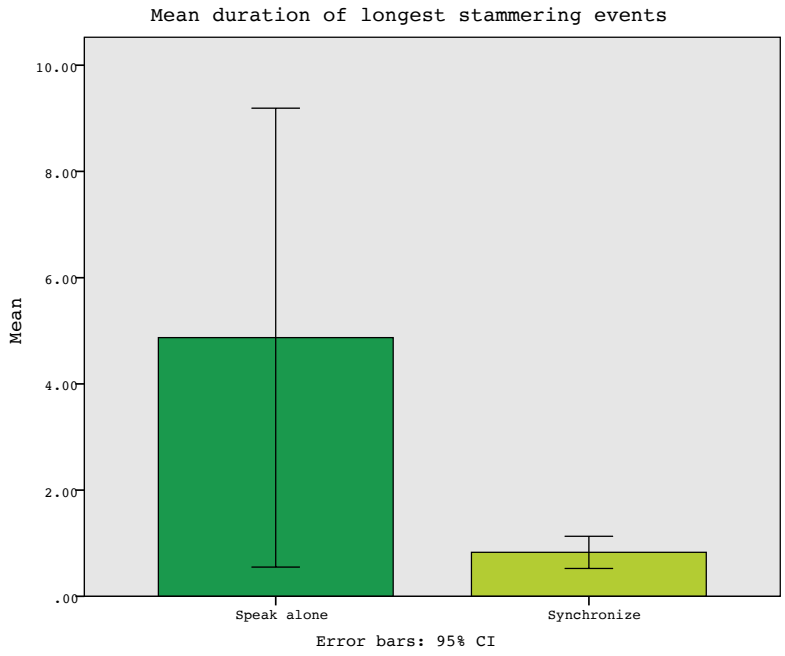


FIGURE 10: MEAN DURATION (S) OF LONGEST STUTTERING INCIDENT, IN QUIET AND DURING SYNCHRONOUS SPEECH

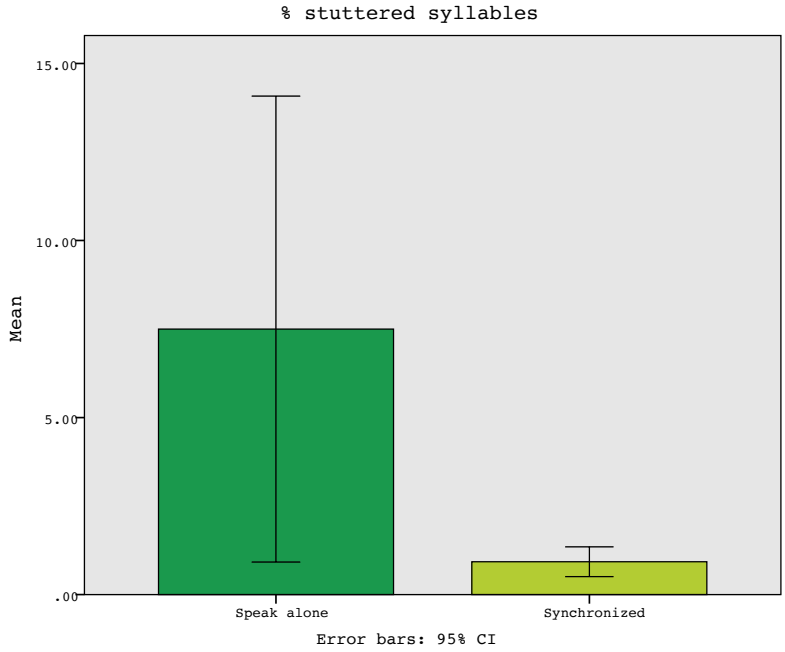


FIGURE 11: MEAN PERCENTAGE OF STUTTERED SYLLABLES IN QUIET AND DURING SYNCHRONOUS SPEECH

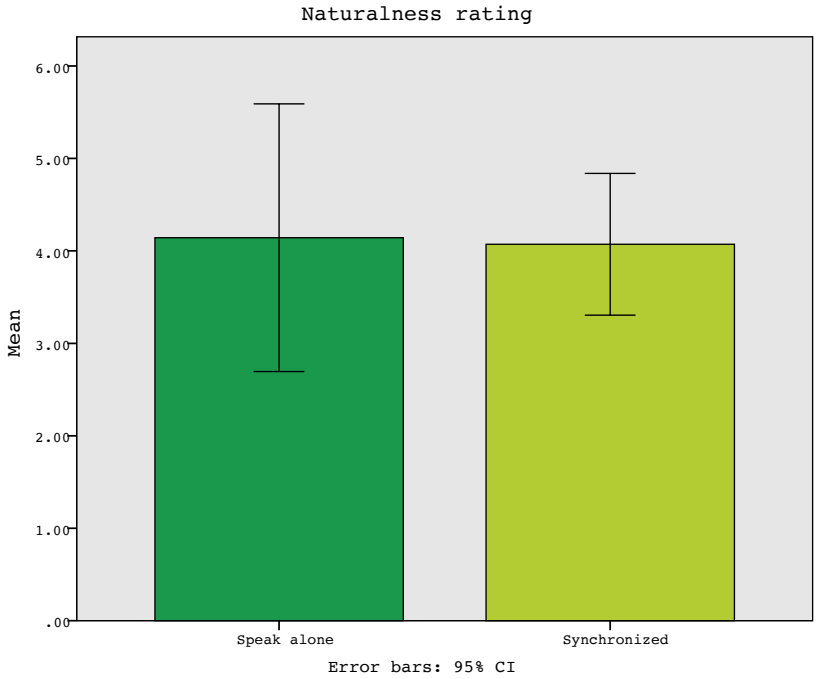


FIGURE 12: MEAN NATURALNESS RATING, IN QUIET AND DURING SYNCHRONOUS SPEECH

SPEECH DATA IN THE SCANNER

For every trial in each of the four speaking conditions, the following parameters were extracted: mean intensity, percentage of stuttered syllables, speech onset and speech rate in fluent syllables per second. Behavioural and acoustic measures were investigated using a linear mixed model with condition as a fixed effect, crossed random effects for subjects and sentences read, and a by-subjects random slope for the effects of condition. This was intended to handle the correlated subject data and address the fact that both subjects and sentences are sampled from a larger population(Barr, Levy, Scheepers, & Tily, 2013; Clark, 1973).

There was a significant effect of speaking condition on speaking rate ($F(3)=26.89$, $q<0.001$) and the percentage of stuttered syllables ($F(3)=13.63$, $p<0.001$); speakers produced fewer fluent syllables per second and stuttered significantly more in the SpeakAlone condition than in other conditions. There was a significant effect of speaking condition on speech intensity ($F(3)=23.62$, $q<0.001$), owing to significantly higher vocal intensity in the SpeakNoise condition than in all other conditions. There was no significant difference in time of speaking onset between conditions ($F(3)=0.944$, $q=0.42$).

A paired-samples t-test was carried out to assess the difference in the degree of synchrony between the participant and the experimenter when the experimenter was live compared to when the participant synchronised with a recording. On average, participants were more closely synchronised with their partner in the SyncLive condition (mean asynchrony score= 0.91, s.d.=0.57) than during SyncRec (mean = 9.36, s.d.=32.53). The mean difference in asynchrony score between conditions was -8.45 (95% confidence intervals -12.9: -4.10), which was statistically significant ($t(204)=-3.75$, $p<0.001$, $d=0.37$).

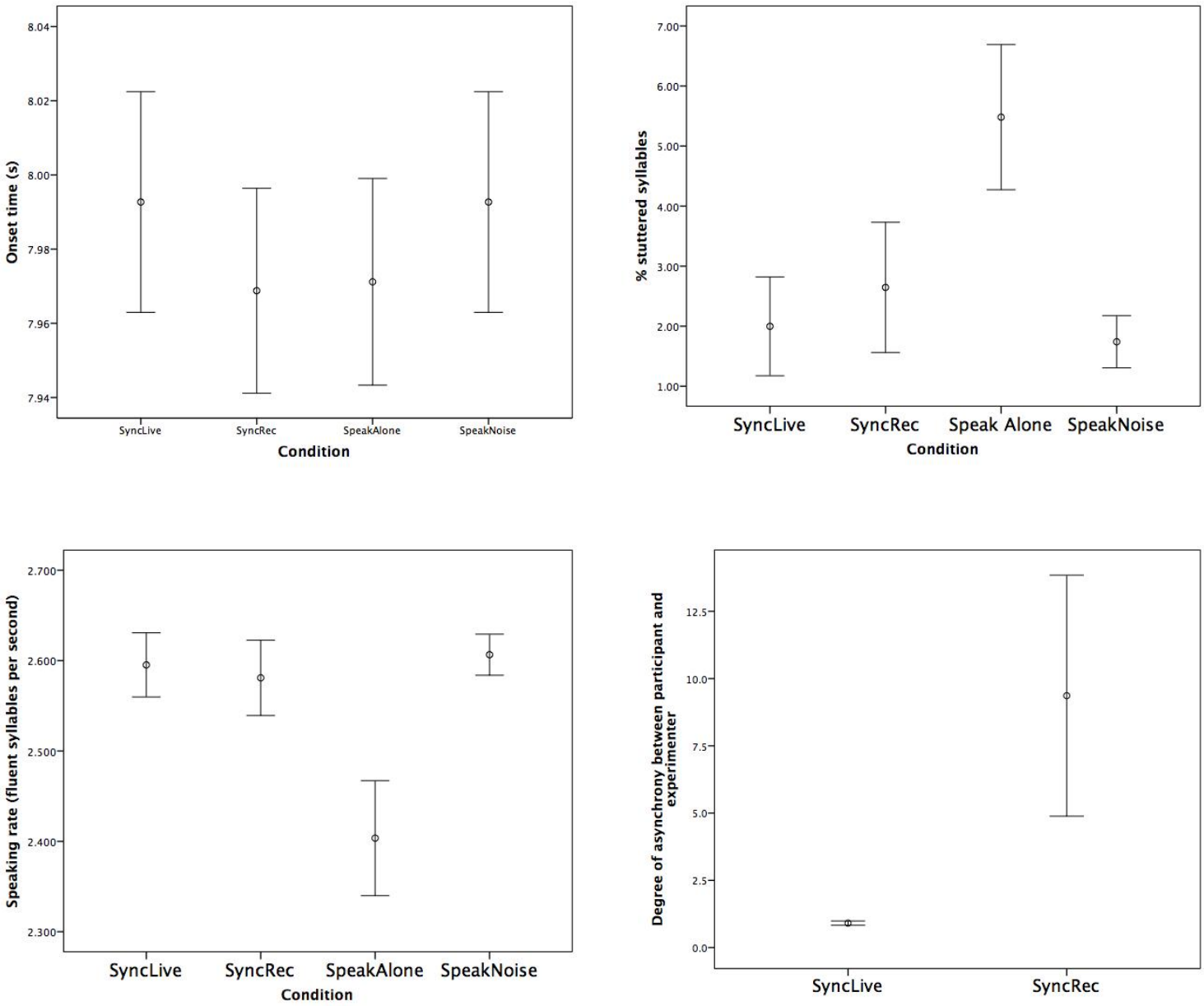


FIGURE 13: IN-SCANNER BEHAVIOURAL DIFFERENCES BETWEEN THE DIFFERENT SPEAKING CONDITIONS (ERROR BARS INDICATING 95% CONFIDENCE INTERVALS). CLOCKWISE FROM TOP LEFT: ONSET TIME IN SECONDS, PERCENTAGE OF STUTTERED SYLLABLES, ASYNCHRONY BETWEEN PARTICIPANT AND EXPERIMENTER AND SPEAKING RATE IN FLUENT SYLLABLES PER SECOND.

UNIVARIATE FMRI

Comparison of the SyncLive and SyncRec conditions showed no differences in neural response, so the two were conflated into one 'synchrony' condition. An ANOVA examining areas of the brain where there were significant differences between one or more of the experimental conditions (Listen, SpeakAlone, SpeakNoise and Synchronize, with ReadSilently as an implicit baseline) revealed widespread activation in bilateral superior temporal cortices extending to postcentral gyri, and cerebellum including bilateral Lobule VI and cerebellar vermis. Additional, smaller clusters were seen in the basal ganglia including thalamus, and parietal, occipital and frontal cortex. To investigate differences between conditions, mean beta values were extracted from selected peak voxels and analysed in SPSS (IBM). As this involved running multiple tests, p-values were FDR corrected for multiple comparisons using the method described by Benjamini and Hochberg (1995), and are reported as corrected q-values.

Peak beta values at [-58 -18 10] in the left STG and [60 -22 0] in the right hemisphere were compared using a repeated measures ANOVA with Hemisphere and Condition as factors. Assumptions of sphericity were met for the Hemisphere factor (as it has only two levels) and for Condition (non-significant Mauchly's W, χ^2 (5)=3.14, p=0.68), but were violated for the interaction between Condition and Hemisphere (significant Mauchly's W, $\chi 2$ (5)=14.31, p= 0.015), so the Greenhouse-Geisser correction for degrees of freedom was applied (ε=0.49). The F-test revealed a main effect of Condition (F(3,24)= 11.5, q<0.001, η_p^2=0.59) and of Hemisphere (F(1,8)= 30.8, q=0.007, η_p^2=0.79) but no significant Condition*Hemisphere interaction (F(1.46,11.66)=2.55, q=0.84). Sidak-corrected posthoc t-tests investigating the main effects of Condition and Hemisphere

showed that, bilaterally, responses were significantly greater in the Synchrony condition than the during the other three tasks ($p>0.05$), with no other significant differences between conditions. Responses in this region were significantly greater in the left than the right hemisphere ($p= 0.001$).

A second ANOVA looked at effects of Condition and Hemisphere in bilateral postcentral gyri at [-46 -12 36] and [54 -8 38]. Mauchly's test showed that the assumption of sphericity was violated for the main effect of Condition (χ^2 (5)=34.1, $p<0.001$) and the interaction between Condition and Hemisphere (χ^2 (5)= 13.7, $p=0.019$), so the Greenhouse-Geisser correction was applied. There was a significant main effect of Condition ($F(1.19, 9.5)= 66.6$, $q<0.001$) but no effect of Hemisphere ($F(1,8)= 2.25$, $q=1.204$) or a significant Condition*Hemisphere interaction ($F(1.38, 11.06)=6.05$, $q=0.168$). Sidak-corrected post-hoc t-tests found that the main effect of Condition was attributable to significantly greater BOLD responses in the three speaking conditions (SpeakNoise, SpeakAlone and Synchronize) than during Listen. There were no other significant differences between conditions.

Two one-way repeated measures ANOVAs investigated neural responses to Condition in the cerebellum. The first looked at responses in the left cerebellum at peak [-14 -64 -22]. The data did not meet the assumption of sphericity (Mauchly's W, χ^2 (5)=22.5, $p<0.001$) so the Greenhouse-Geisser estimates of degrees of freedom were used ($\varepsilon=0.41$). The F-test showed a significant main effect of Condition ($F(1.22,9.78)=32.43$, $q<0.001$, $\eta_p^2=0.80$). Sidak-corrected post-hoc tests showed that activation in the Listen condition was significantly lower than in all other conditions ($p<0.004$). The second F-test investigated the effect of Condition in the cerebellar vermis at peak [-2 -44 -20].

Mauchly's test was significant, indicating non-sphericity (χ^2 (5)=12.59, p=0.029) so the Greenhouse-Geisser correction was applied (ε=0.51). There was a significant effect of Condition (F(1.5,12.1)=7.11, q=0.013, η_p^2=0.47), which post-hoc Sidak corrected t-tests demonstrated was owing to a significantly lower response in the Listen condition than in SpeakNoise (p=0.028).

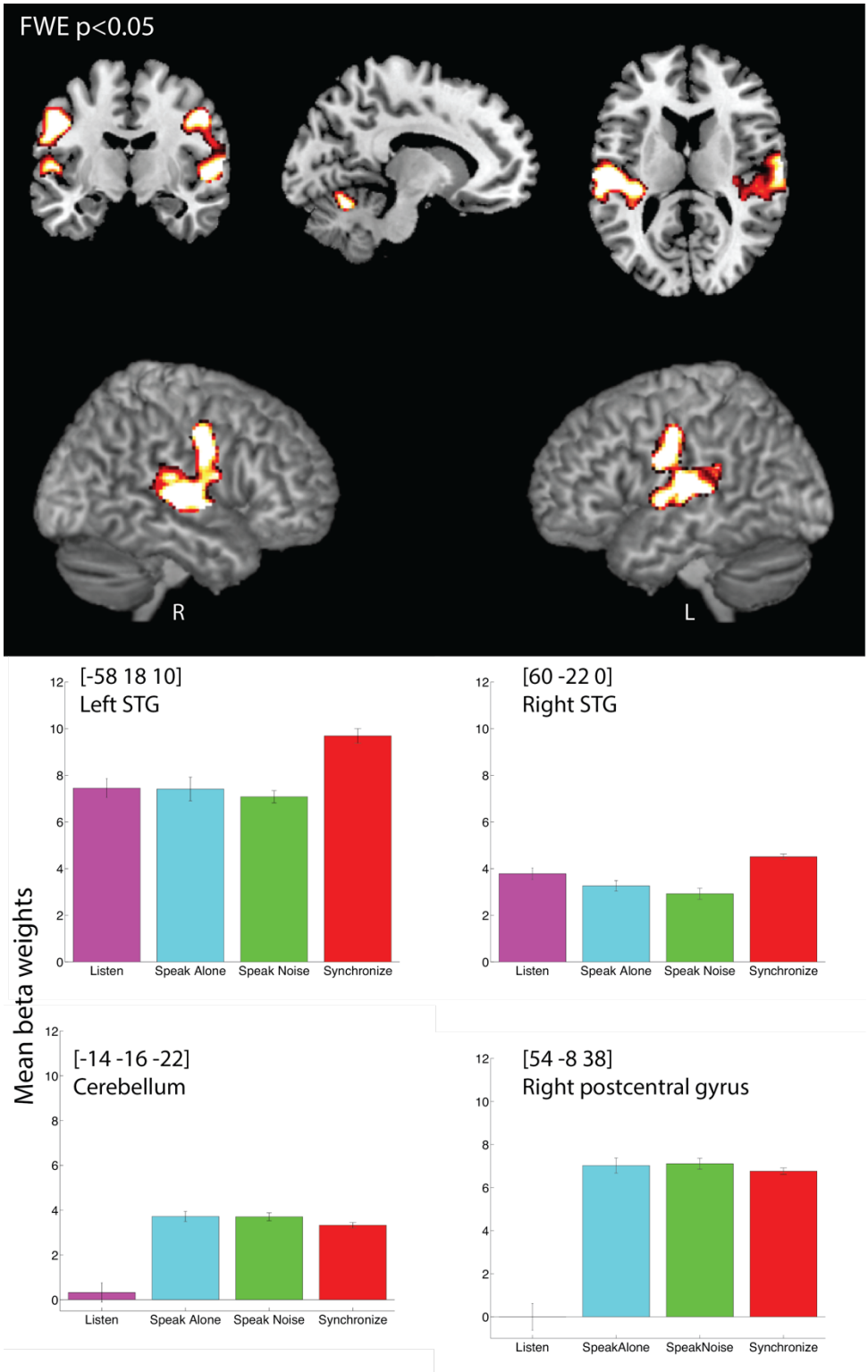


FIGURE 14: MEAN BETA WEIGHTS AT PEAK VOXEL CO-ORDINATES REVEALED BY AN ANOVA COMPARING LISTEN, SPEAKALONE, SPEAKNOISE AND SYNCHRONIZE, WITH THE REST CONDITION AS A BASELINE. CORRECTED FOR MULTIPLE COMPARISONS AT FWE P<0.05 WITH EXTENT THRESHOLD OF 10 VOXELS.

TABLE 8: PEAK VOXEL CO-ORDINATES IN REGIONS MODULATED BY THE DIFFERENT TASKS, REVEALED BY A ONE-WAY ANOVA CONTRASTING LISTEN, SPEAKALONE, SPEAKNOISE AND SYNCHRONISE

Anatomy	Voxels (k)	Z-score	x	y	z
Left STG	6334	Inf	-58	-18	10
Left postcentral gyrus		Inf	-46	-12	36
Left postcentral gyrus		Inf	-58	-4	24
Right postcentral gyrus	6287	Inf	54	-8	38
Right STG		Inf	60	-22	0
Right rolandic operculum		Inf	42	-30	18
Left cerebellum	5559	Inf	-14	-64	-22
Left cerebellum (Lobule VI)		Inf	-22	-60	-22
Right cerebellum (Lobule VI)		7.79	14	-66	-20
Cerebellar vermis	45	6.41	-2	-44	-20
Left Superior parietal lobule	229	6.35	-38	-54	62
Left inferior parietal lobule		5.73	-56	-42	52
Left inferior parietal lobule		5.63	-42	-60	56
Right middle occipital gyrus	39	6.07	40	-86	18
Left angular gyrus	269	5.98	-48	-76	24
Left middle occipital gyrus		5.54	-40	-82	28
Left angular gyrus		5.39	-42	-56	22
	35	5.92	4	20	12
Right thalamus	38	5.88	16	-16	8
Thalamus (prefrontal)		5.61	18	-8	8
	48	5.58	38	2	6
	26	5.56	-30	28	18
		5.44	-24	22	18
Right precuneus	61	5.53	8	-78	46
Left cuneus		4.95	2	-84	34
Left precuneus		4.85	-2	-78	42
Area 4p	23	5.25	20	-26	58

Left inferior occipital gyrus	11	5.22	-54	-62	-14
Left posterior-medial frontal	32	5.04	0	2	68
Left posterior-medial frontal		4.97	-4	2	60

REGIONS ASSOCIATED WITH STUTTERING SEVERITY

A multiple regression was carried out with the percentage of stuttered syllables on each trial as a regressor. This revealed correlations between neural activation and stuttering frequency regardless of experimental condition. Stuttering severity was associated with the BOLD response in network of areas in the frontal cortex with clusters in bilateral inferior frontal gyri, and in cerebellum including cerebellar vermis. Smaller clusters were also seen in bilateral precentral and postcentral gyri, precuneus and right superior parietal cortex.

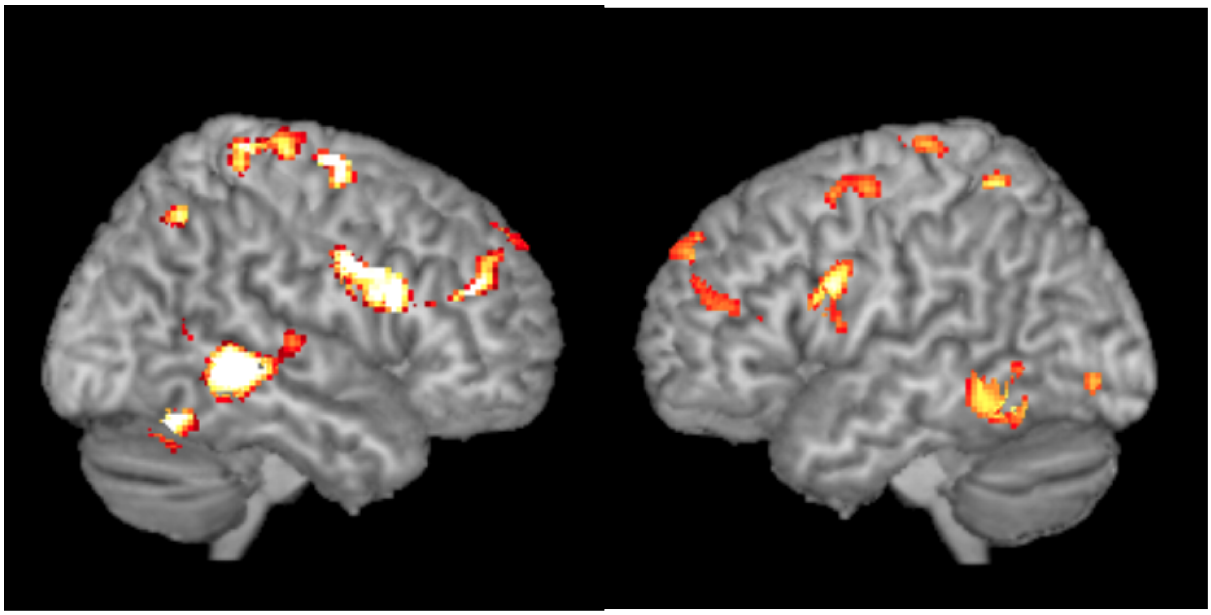


FIGURE 15: ACTIVATION POSITIVELY CORRELATED WITH INCREASED PERCENTAGE OF SYLLABLES STUTTERED, REVEALED BY A MULTIPLE REGRESSION ANALYSIS. THRESHOLDED AT FWE P<0.05 WITH EXTENT THRESHOLD OF 50 VOXELS

TABLE 9: PEAK VOXEL CO-ORDINATES REVEALED BY A MULTIPLE REGRESSION ANALYSIS CORRELATING BOLD RESPONSE WITH PERCENTAGE OF STUTTERED SYLLABLES

Anatomy	Voxels (k)	Z score	x	y	z {mm}
Left IFG- Area 44	762	6.34	-68	10	20
		6.12	-78	-8	14
		5.84	-76	-4	2
	963	6.2	-70	-62	-14
		6.07	-58	-80	-12
		6.05	-74	-54	-8
	774	6.12	66	2	32
Right IFG (pars opercularis)		5.93	60	12	30
Right IFG (pars triangularis)		5.92	56	16	24
Right cerebellum (Crus 1)	1148	6.08	56	-64	-24
		6.02	74	-42	-4
		5.79	56	-86	-6
	189	6.06	-42	56	28
		5.05	-54	44	18
Left middle frontal gyrus		4.86	-36	46	18
Left superior parietal	55	5.79	-28	-52	62

Right Superior frontal gyrus	111	5.71	28	-8	68
Right superior frontal gyrus		5.43	36	-4	64
Right middle temporal gyrus	63	5.63	48	-52	6
Left cerebellum (IV-V)	469	5.56	-4	-50	-2
Cerebellar vermis		5.51	-2	-60	-6
cerebellar vermis		5.33	6	-50	-4
Right postcentral gyrus	215	5.46	22	-36	74
Right precentral gyrus		5.2	28	-22	72
Right postcentral gyrus		5.11	32	-38	68
Right superior medial gyrus	286	5.45	6	52	40
Right superior frontal gyrus		5.35	14	50	32
Left superior medial gyrus		5.21	-8	56	36
Right superior parietal gyrus	83	5.29	14	-56	64
Right precuneus		4.68	4	-62	56
Left postcentral gyrus	71	5.26	-20	-26	76

Right angular gyrus	72	5.24	38	-60	48
Left precentral gyrus	66	5.14	-38	-6	60

INDEPENDENT COMPONENT ANALYSIS

Nine non-artefactual independent components were identified. Temporal sorting was carried out to correlate each of the component timecourses with the SPM design matrix- this gave a correlation coefficient indicating how strongly task-related each component was. Next, probabilistic labels were applied by correlating the spatial map with templates included in the GIFT toolbox. These are displayed in table N below. Four of the components (C3, C111, C15 and C16) were identified as default mode networks by GIFT, confirmed by visual inspection. FDR-corrected repeated measures ANOVAs were carried out on the five remaining components to assess how each network was modulated by the five experimental conditions. The component spatial maps were corrected for multiple comparisons using voxelwise FDR $p<0.05$ and a cluster extent threshold of 50 voxels; the resulting activation maps are shown below.

TABLE 10: COMPONENTS IDENTIFIED BY GROUP ICA ANALYSIS WITH THEIR PROBABILISTIC ANATOMICAL NETWORK CORRELATES AS ASSIGNED BY THE GIFT TOOLBOX

Component ID	Label	Correlation with predefined network (R)
C1	Intraparietal sulcus/frontal eye fields	0.1928
C3	PCC/MPFC (Dorsal default mode network)	0.2942
C5	Auditory network	0.3892
C6	Higher visual network	0.2803
C8	Precuneus network	0.2123
C10	Left DLPFC/Parietal (Left Executive Control Network)	0.1341
C11	PCC/MPFC (Dorsal default mode network)	0.2641
C15	Retrosplenial cortex/medial temporal lobe (Ventral default mode network)	0.3348
C16	PCC/MPFC (Dorsal default mode network)	0.3348f

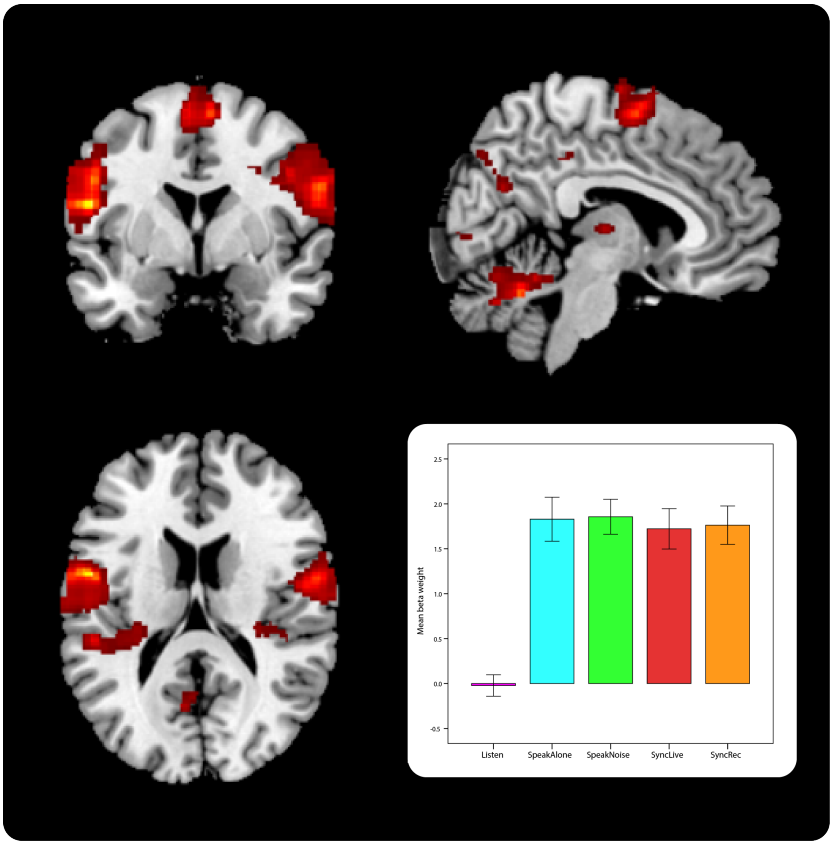


FIGURE 16: COMPONENT 1

TABLE 11: CO-ORDINATES AND PROBABILISTIC ANATOMICAL LABELS FOR PEAKS WITHIN COMPONENT 1

Anatomy	Voxels (k)	Z-score	x	y	z {mm}
Right postcentral gyrus	2480	5.64	60	-6	28
Right postcentral gyrus		5.35	60	-10	38
Right postcentral gyrus		4.74	60	-2	20
Left Postcentral gyrus	2940	5.52	-60	-10	26
Left precentral gyrus		5.39	-52	0	18
Left postcentral gyrus		5.15	-58	-16	32
Left cuneus	715	5.18	-4	-66	24
Left precuneus		4.39	-8	-60	34
Left precuneus		3.89	-8	-70	32
Cerebellar vermis	1910	4.83	4	-56	-24
Cerebellar vermis		4.51	-2	-70	-16
Cerebellar vermis		4.42	4	-68	-30
	437	4.7	-28	-24	12
Left superior temporal gyrus		4.05	-52	-32	18
Left supramarginal gyrus		3.3	-44	-38	24

Right thalamus	159	4.61	8	-14	6
Thalamus (premotor)		3.43	16	-16	-4
Thalamus (premotor)		3.1	20	-12	4
Right posterior-medial frontal	670	4.56	6	0	62
Left posterior-medial frontal		4.14	-4	0	62
Right posterior-medial frontal		3.73	6	-8	74
Left thalamus	176	4.4	-14	-20	6
Left thalamus		4.39	-20	-22	0
Left inferior parietal lobule	102	4.25	-38	-50	38
	71	4.05	-10	20	26
		3.13	-10	10	30
	113	3.89	32	-30	22
Right rolandic operculum		3.1	46	-32	20
Right precentral gyrus	113	3.54	22	-28	62
Right calcarine gyrus	65	3.45	10	-80	4
Right lingual gyrus		3.28	12	-84	-8
Right lingual gyrus		3.13	20	-84	-12
Left precentral gyrus	57	2.97	-22	-28	56

Component 1 was strongly task-related (r=0.73) and was labelled as a visuospatial network (r=0.19). Although there were some clusters in occipital and frontal cortex, this component included large clusters in bilateral postcentral and precentral gyri, left superior temporal gyrus, and subcortical structures including left cuneus and precuneus, and cerebellar vermis; a full list of peaks and their anatomical labels is given in Table 12 above. It was considered more likely, therefore, that this component reflected sensorimotor processing. A repeated-measures ANOVA was carried out looking at differences between the experimental conditions. Mauchly's W was significant ($\chi 2$ (9)=67.2, p<0.001), indicating that the assumption of sphericity had been violated, so the Greenhouse-Geisser correction for degrees of freedom was used (ε=0.48). There was a significant effect of condition (F(1.92, 49.9)=175.1, q<0.001, η_p^2=0.87) which Sidak-corrected post-hoc tests revealed was owing to lower beta values in the Listen condition

than all other conditions ($p<0.001$) and in SyncRec than SpeakNoise ($p=0.01$). This suggests that the network was involved in articulation and speech production; the increased response to SpeakNoise compared to SyncRec could be explained by subtle differences in articulation related to increasing vocal intensity.

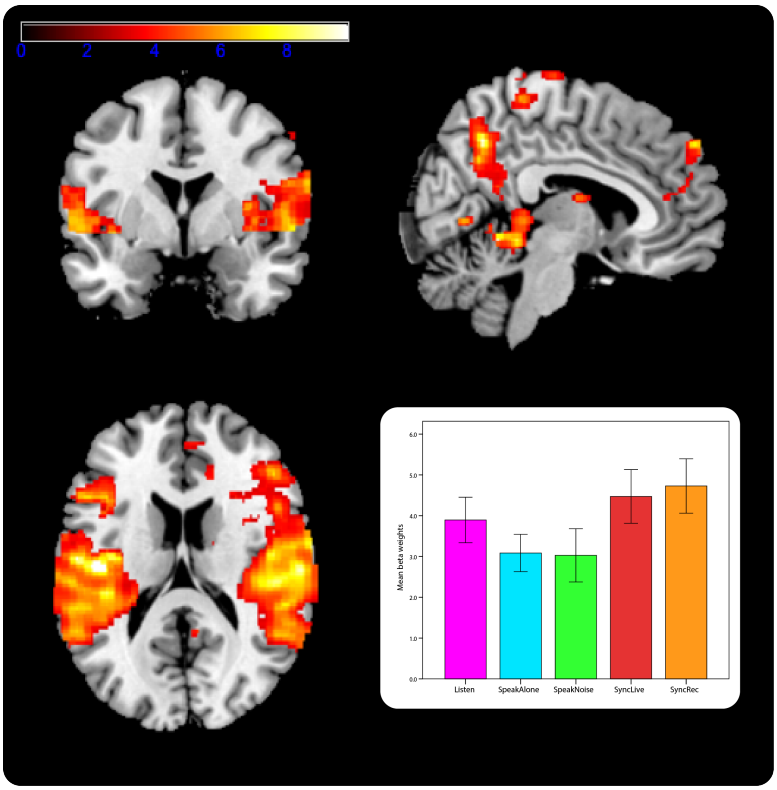


FIGURE 17: COMPONENT 5

TABLE 12: CO-ORDINATES AND PROBABILISTIC ANATOMICAL LABELS FOR PEAKS WITHIN COMPONENT 5

Anatomy	Voxels (k)	Z-score	x	y	z {mm}
Thalamus	7132	5.37	-30	-22	2
Left rolandic operculum		5.13	-44	-16	18
Left insula		4.92	-36	-16	4
Right superior temporal gyrus	12772	5.14	64	-18	8
Right precentral gyrus		4.78	46	-12	56
Right superior temporal gyrus		4.77	54	-32	4
Left lingual gyrus	1474	4.94	-18	-56	2
Right cerebellum (IV)		4.88	8	-50	-8
Cerebellar vermis		4.84	0	-44	-2

Right precuneus	1181	4.83	8	-58	44
		4.37	-14	-48	30
Right precuneus		4.25	2	-60	36
BA 5	115	4.24	-14	-46	60
Left precuneus		3.16	-6	-44	60
Left IFG (pars triangularis)	490	4.24	-50	20	16
Left IFG (pars opercularis)		3.38	-42	10	16
Left IFG (pars triangularis)		3.32	-52	22	8
	262	4.18	8	58	42
		3.31	16	32	16
Left ACC		3.03	2	46	16
Right thalamus	63	4.11	10	-8	14
Left posterior-medial frontal	119	4.09	-10	-16	52
Left posterior-medial frontal		2.65	0	-20	52
	79	3.94	-24	20	10
Left ACC	60	3.84	0	24	22
Left lingual gyrus	54	3.82	-22	-68	-12
Left cerebellum (VI)		2.45	-28	-70	-18
Right cerebellum (VI)	243	3.8	14	-72	-18
Right lingual gyrus		3.47	16	-58	-10
Right cerebellum (Crus 1)		3.33	14	-76	-32
Left Cerebellum (Crus 1)	362	3.72	-24	-80	-30
Left Cerebellum (Crus 1)		2.97	-8	-72	-30
Left cerebellum (VI)		2.89	-16	-64	-28
Right lingual gyrus	69	3.59	6	-68	2
Left lingual gyrus		2.57	-4	-66	2
Left middle occipital gyrus	70	3.28	-54	-72	0
Left inferior occipital gyrus		3.06	-50	-72	-8
Left inferior temporal gyrus		3.03	-52	-58	-18
	121	3.28	-26	-30	46
Left postcentral gyrus		3.07	-40	-36	62
Left postcentral gyrus		2.9	-34	-32	56
Right cerebellum (IV)	109	3.18	28	-46	-20
Right fusiform gyrus		2.98	42	-48	-22
Right cerebellum (VI)		2.83	34	-56	-24
	72	3.11	-62	-30	46
Left inferior parietal lobule		2.98	-54	-30	50

Component 5, identified by GIFT as an auditory network (r=0.39), included clusters in the right STG and left IFG, as well as cerebellum and precuneus. The component was moderately task-related (r=0.56). Mauchly's test demonstrated that the data did not meet the assumption of sphericity ($\chi2(9)=18.7$, $p=0.03$) and the Huynh-Feldt correction for degrees of freedom was used ($\varepsilon=0.92$). A repeated-measures ANOVA demonstrated that there was a significant effect of condition ($F(3.68, 95.5)=27.29$, $q<0.001$, $\eta_p^2=0.51$). Sidak-corrected post-hoc tests showed a complex pattern of differences between conditions, with significantly lower beta values in the SpeakAlone and SpeakNoise conditions than in SyncLive and SyncRec ($p<0.001$), and in Listen than SyncRec. SyncLive was not significantly different to SyncRec ($p= 0.33$) or to Listen ($p=0.08$), and there was also no significant difference between SpeakAlone and SpeakNoise ($p= 1.0$).

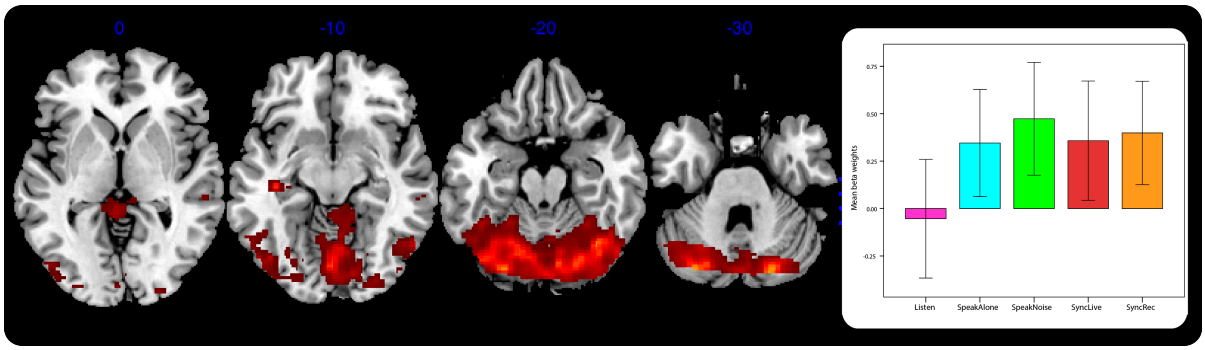


FIGURE 18: COMPONENT 6

TABLE 13: CO-ORDINATES AND PROBABILISTIC ANATOMICAL LABELS FOR PEAKS WITHIN COMPONENT 6

Anatomy	Voxels (k)	Z-score	x	y	z {mm}
Left cerebellum (Crus 1)	8415	5.81	-28	-80	-28
Right cerebellum (Crus 1)		5.78	22	-78	-26
Right cerebellum (VI)		5.4	12	-78	-22
	118	4.82	-40	-24	-8
Thalamus		4.67	-30	-22	-4
Right precuneus	201	3.99	4	-66	48
Right precuneus		3.25	4	-56	66
Right precuneus		3.15	4	-52	58
Right middle temporal gyrus	110	3.73	66	-30	-8
Right superior temporal gyrus		3.25	58	-20	-4
Right middle temporal gyrus		3.17	56	-30	0
Right ACC	69	3.34	4	32	18

Component 6 was weakly task-related (r=0.23) and contained peaks in cerebellum bilaterally, right precuneus, superior and middle temporal gyri, and anterior cingulate cortex. It was weakly correlated with the GIFT predefined higher visual network (r=0.28). Mauchly's test indicated that the assumption of sphericity was met for this data ($\chi 2(9)=16.64$, p=0.055). There was a significant main effect of condition ($F(4,104)=4.95$, q=0.01, $\eta_p^2=0.16$) which post-hoc t-tests (Sidak corrected for multiple comparisons)

revealed was attributable to lower beta values in the Listen condition than during SyncRec (p=0.019); there were no other significant differences between conditions.

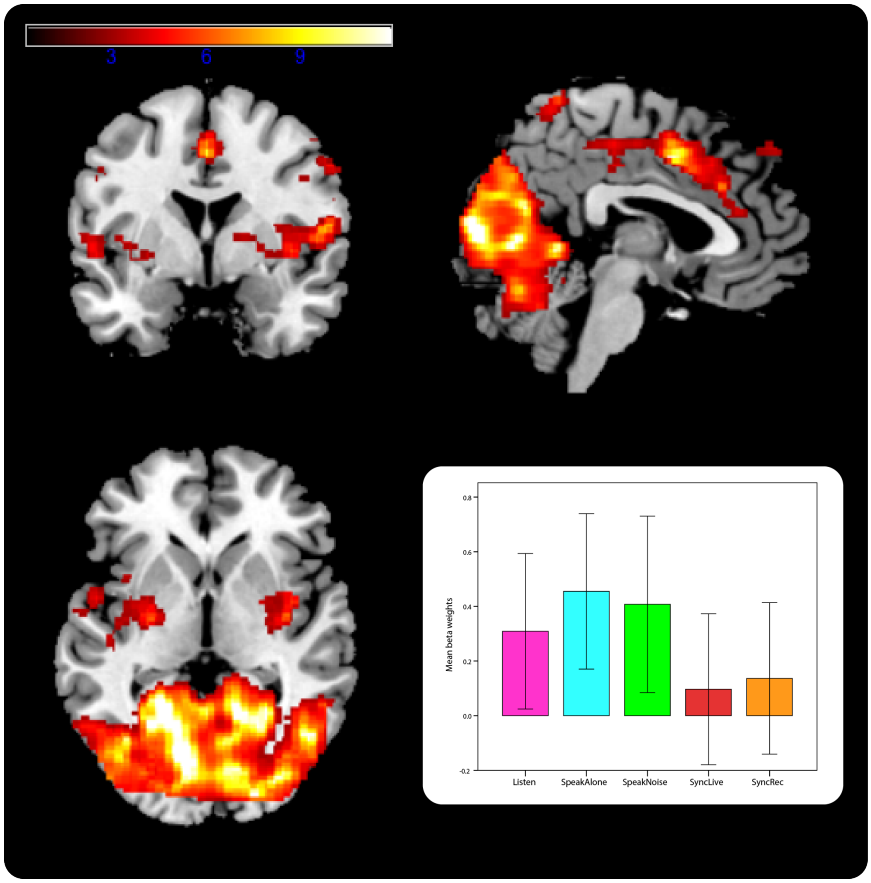


FIGURE 19: COMPONENT 8

TABLE 14: CO-ORDINATES AND PROBABILISTIC ANATOMICAL LABELS FOR PEAKS WITHIN COMPONENT 8

Anatomy	Voxels (k)	x	y	z {mm}
Left lingual gyrus	26943	-12	-74	-2
Right calcarine gyrus		28	-56	4
Left precuneus		-20	-50	0
Left MCC	1260	0	4	44
Right MCC		10	8	44
Right MCC		2	26	30
Left rolandic operculum	460	-46	-24	22
Left insula		-34	-22	18

Left rolandic operculum		-52	-16	14
Right supramarginal gyrus	3177	60	-24	24
Rigt supramarginal gyrus		66	-20	32
Right precentral gyrus		60	-8	42
Left precentral gyrus	430	-54	10	34
Left middle frontal gyrus		-44	24	38
Left middle frontal gyrus		-42	12	50
Right precuneus	609	10	-54	70
Right superior parietal lobule		18	-54	64
Left precuneus		-2	-56	64
Left postcentral gyrus	84	-24	-34	68
Left putamen	439	-30	-4	-6
Left pallidum		-26	-10	-2
		-32	12	0
Right superior frontal gyrus	677	16	50	48
Right superior frontal gyrus		18	40	52
Left superior frontal gyrus		-26	46	42
Right postcentral gyrus	56	30	-30	72
Right superior frontal gyrus	124	18	-12	70
Right precentral gyrus		30	-18	70
Right posterior-medial frontal		8	-6	76
Left inferior parietal lobule	177	-40	-28	36
Left supramarginal gyrus		-58	-28	32
Left supramarginal gyrus		-50	-26	34
Right thalamus	134	8	-18	14
		20	-32	18
Right thalamus		16	-24	12
Left caudate nucleus	52	-10	14	6
Left IFG (pars triangularis)	66	-54	30	10
Left IFG (pars triangularis)		-48	30	16
Left middle frontal gyrus		-42	40	20

Component 8 was weakly modulated by task (r=0.18) and was correlated with GIFT's map of the precuneus network at r=0.21. Within this network, there were peaks in precuneus, left caudate nucleus, putamen and pallidum, and right thalamus. At the cortical level, the network included clusters in bilateral occipital cortex, middle cingulate cortex, and superior and middle frontal cortex. Mauchly's test was significant

($\chi 2(9)=42.5$, $p<0.001$), indicating that the assumption of sphericity was violated, so the Greenhouse-Geisser correction was applied ($\varepsilon=0.54$). A repeated measures F-test revealed a significant main effect of condition ($F(2.2,56.6)=3.67$, $q=0.004$, $\eta_p^2=0.124$). This was followed up by Sidak-corrected post-hoc t-tests which showed that the network was significantly more modulated by SpeakAlone and SpeakNoise than by SyncLive and SyncRec ($p<0.03$); there were no other significant differences between conditions.

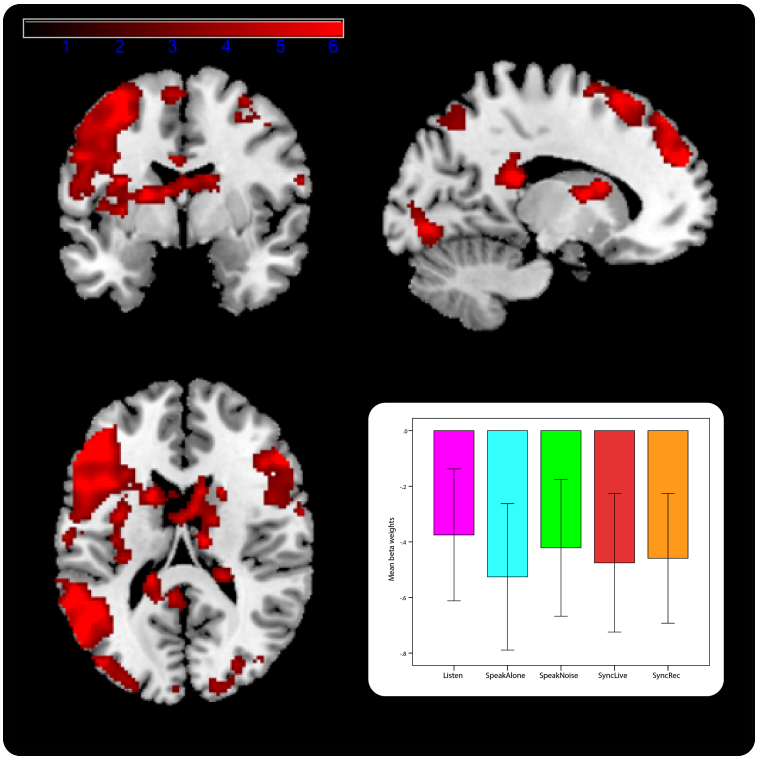


FIGURE 20: COMPONENT 10

TABLE 15: CO-ORDINATES AND PROBABILISTIC ANATOMICAL LABELS FOR PEAKS WITHIN COMPONENT 10

Anatomy	Voxels (k)	Z-score	x	y	z {mm}
Left inferior parietal lobule	25363	5.78	-48	-58	44
Left IFG (pars triangularis)		5.46	-50	30	10
Left middle frontal gyrus		5.16	-40	22	34
Right IFG (pars triangularis)	2013	5.34	54	28	20
Right IFG (pars triangularis)		5	50	26	28
Right IFG (pars opercularis)		4.61	46	14	32
Right inferior parietal lobule	1866	4.53	46	-34	52
Right inferior parietal lobule		4.31	38	-46	54
Right inferior parietal lobule		4	40	-48	42
Left ACC	71	4.41	0	6	26
	906	4.34	-16	-36	18
Left MCC		4.21	-8	-28	36
Left PCC		3.96	-2	-42	20
Right cerebellum (VI)	2429	4.18	24	-80	-18
Right middle occipital gyrus		4.1	34	-92	6
Right middle occipital gyrus		4.09	28	-92	12
Right middle frontal gyrus	238	3.97	38	6	52

Right precentral gyrus		3.76	44	2	48
Right middle frontal gyrus		3.33	36	2	60
Right paracentral lobule	107	3.42	8	-20	78
Right paracentral lobule		3.31	10	-30	74
Left cerebellum (Crus 1)	52	3.37	-34	-72	-30
Left cerebellum (VI)		2.8	-24	-66	-30
	141	3.3	18	-32	18
		3.13	30	-24	6
		2.75	32	-32	10
Right middle temporal gyrus	369	3.26	56	-46	0
Right superior temporal gyrus		3.24	56	-28	10
Right Heschl's gyrus		3.12	48	-22	8
Left cuneus	82	3.18	-2	-94	20
Right cuneus		2.68	4	-72	20
Left cuneus		2.41	2	-84	20

Component 10 covered a spread of regions bilaterally in middle and inferior frontal cortex, the inferior parietal lobule, cuneus, left cingulate cortex and right middle and superior temporal gyrus. It was weakly modulated by task (r=0.14) and was weakly correlated with the left executive control network (r=0.13). Mauchly's test indicated non-sphericity ($\chi 2(9)=18.7$, p=0.28) so the Huynh-Feldt correction was applied ($\varepsilon=0.86$). A repeated measures ANOVA revealed no significant effect of condition ($F(3.4,89.4)=0.63$, $q=0.62$, $\eta_p^2=0.024$).

6.6. CONCLUSIONS

Behavioural performance both in and out of the scanner confirmed that synchronised speech is extremely effective at inducing fluency in people who stutter, regardless of stuttering severity. Subjects' speech contained fewer stuttering incidents, and the incidents were of shorter duration, when they spoke chorally compared to speaking alone. Masking noise also reduced the percentage of syllables stuttered and, contrary to expectations, there were no significant differences in measures of stuttering severity when participants spoke in noise compared to when they synchronised with a partner, suggesting that in this experiment, both altered feedback techniques were equally effective at inducing fluency. The two synchronous speech conditions, SyncLive and SyncRec, had similar effects on fluency, but analysis of the recordings showed that participants synchronized more effectively when they spoke with a live experimenter than when they were synchronizing with a recording, confirming the effect identified by Jasmin et al (2016). However, the functional analysis failed to find a neural distinction between the two synchrony conditions. Jasmin et al (2016) found that synchronising with a recording was associated with suppression in temporal cortex relative to listening to sounds, while synchronising with a live partner resulted in a release from this suppression. Here, a univariate analysis showed that responses in the STG bilaterally were the same for speaking alone, listening to speech, and speaking in masking noise, but significantly greater when participants spoke in synchrony either with a live partner or with a recording. Note however that we did not directly replicate Jasmin et al's analysis (a region of interest analysis limited to the right temporal pole), potentially explaining the difference in results. Our results suggest that PWS do not display a speaking-induced

suppression response, which may support the theory that stuttering arises from an over-reliance on auditory feedback. However, if this is the case and disfluency is related to STG over-activation then it is unclear why synchronous speech, which induces fluency, should be associated with an increase in STG activation. It is possible that the response to choral speech in the STG could reflect a preferential response to informational masking, such as that found in the study described in chapter 5 (Meekings et al., 2016). This is especially plausible as, to synchronize effectively, it was necessary for participants to attend closely to the speech of their conversational partner.

We also saw activation in the cerebellum associated with the different speech production conditions. Further analysis demonstrated that activity in several cerebellar regions was positively correlated with stuttering severity. This included the cerebellar vermis, which has previously been implicated in dysfluency (Brown et al., 2005) supporting Budde et al's (2014) finding that activation in cerebellar vermis is associated with state stuttering (though c.f. Belyk et al, 2014). The regression analysis also found activation in motor and premotor areas consistent with the fact that stuttering involves greater movement of the articulators. A possible future analysis integrating neural and behavioural data could look at activation associated with natural fluency (that is, fluency in the SpeakAlone condition) versus activation associated with induced fluency (fluency in the SpeakNoise and choral speech conditions), as suggested by Budde et al. (2014).

More complex patterns of activation were revealed by the independent component analysis. Of particular interest were Component 8, which involved the basal ganglia, and Component 5, an auditory network which included right STG, left IFG and cerebellum. Component 8 was modulated more by SpeakNoise and SpeakAlone than the two

synchrony conditions, while Component 5 was modulated more by SyncRec and SyncLive than by SpeakNoise and SpeakAlone (and more by SyncRec than Listen). Basal ganglia infarcts have previously been associated with stuttering (Alm, 2004; Giraud et al., 2008). Here, activity in the basal ganglia was modulated more by SpeakAlone, in which more stuttered syllables were produced, and SpeakNoise, which is generally considered a less effective fluency inducing technique than synchronized speech. Meanwhile, auditory areas including right STG responded more to the synchrony conditions. This implies clearly dissociable mechanisms for producing synchronized speech compared to other types of speech production. The difference in networks recruited by choral speech compared to masked speech may help explain why masked speech is less reliable at inducing fluency (Garber & Martin, 1974), although it should be noted that no differences in fluency were found between masking noise and synchronous speech in this study.

This experiment was designed to test the theory that stuttering results from an over-reliance on auditory feedback. The evidence on this point is inconclusive. On the one hand, we found over-activation in the STG when participants spoke alone, which appears to support the over-activation hypothesis. However, synchronised speech was associated with an even greater STG response, which is unexpected if the STG supports error monitoring. Based on our previous finding that activation in the STG is modulated by informational masking during speech production (Meekings et al., 2016), as well evidence from the ICA analysis that right STG activation is modulated more by choral speech than masked speech (despite similar levels of auditory 'error'), it is likely that we are seeing evidence for multiple streams of processing in the STG. The results of our other analyses additionally point to a role for the basal ganglia and cerebellum in stuttering,

consistent with previous research (Giraud et al., 2008; Belyk et al., 2014). Activity in these areas was associated with speaking alone and talking over a noise masker, and was correlated with stuttering severity. Since these regions are involved in the timing of self-paced movement, this may provide evidence that stuttering arises from a deficit in movement timing and regulation. However, it should be noted that speech rate did not significantly differ between altered speech conditions, and was significantly slower when participants spoke alone (and stuttered), rather than being positively associated with increased fluency as might be expected.

There were a number of differences between this study and previous research, most notably our finding that PWS did not display a speaking-induced suppression response- that is, the STG was over-active when participants spoke alone, rather than under-active as previously suggested (Brown et al, 2015). Additionally, despite previous studies suggesting that activity in auditory and motor cortex is right-lateralized in PWS, in our sample we found that peak activity was greater in the left hemisphere than in the right.To confirm these results, it would be desirable to recruit a larger sample of people who stutter, as well as a control group- unfortunately, constraints on time and resources meant that this was not possible at the time of testing. It should be noted that there is disagreement even among large-scale meta-analyses about the neural hallmarks of stuttering as a trait or state (Budde et al., 2014; Belyk et al., 2014). Stuttering may be an 'umbrella syndrome' composed of multiple disorders with overlapping symptoms but distinct aetiologies. For example, stuttering can be caused by head injury (Alm, 2000) and there is considerable individual variability within subjects (Wymbs, Ingham, Ingham, Paolini, & Grafton, 2013). In this study we found several unique results. In our sample of

people who stutter, synchronized speech recruits a distinct network of cortical and cerebellar regions that are not modulated by other types of speech production. Additionally, our analysis showed that the STG does not distinguish between hearing sounds, speaking alone, and talking in masking noise, lacking the speaking-induced suppression response seen in typical speakers. However, it responds significantly more to speaking synchronously, potentially reflecting different streams of processing within auditory cortex. Although these results do not provide unequivocal support for the feedback over-reliance hypothesis, they have implications for our understanding of how the STG works and how PWS process speech. Further research with an expanded sample size and control group can confirm our findings and contribute to our understanding of how PWS differ from typical speakers.

CHAPTER 7: CONCLUSIONS

The experiments in this thesis were designed to explore the role of the superior temporal gyrus in auditory feedback control of speech, and specifically to test the assertion that the STG encodes match and mismatch between auditory feedback and auditory targets during speech production

The questions addressed were as follows:

1. Is the STG reliably activated by feedback perturbation?

2. Do lesions involving the STG result in problems with feedback control?

3. Do people with a hypothesised impairment in feedback control display anomalous STG activation?

These questions were intended to provide some insight into the more general question of how central feedback control is to speech production, and what the definition of a feedback mismatch or 'speech error' should be. The following sections review these questions and discuss the novel experimental findings of this thesis.

7.1. Is the STG reliably activated by feedback perturbation?

If activation in the superior temporal gyrus is modulated by the degree of mismatch between auditory targets and feedback, then anything that perturbs feedback and creates a mismatch should be associated with stronger responses in the STG compared to speaking with no perturbation, and listening. Moreover, the amount that feedback is perturbed should correlate with activation in the STG: the greater the mismatch, the greater the activation.

Two studies in this thesis addressed this question. In chapter 3, an ALE meta-analysis of functional imaging studies that had compared neural responses to speaking with feedback perturbation to unperturbed speech found a significant convergence between reported peak co-ordinates in the bilateral posterior STG, with slightly more anterior clusters in the left hemisphere. However, a systematic review accompanying the meta-analysis found that many of the studies included failed to find results at the whole brain level when corrected for multiple comparisons, indicating a relatively weak effect. Additionally, more than half of the studies included did not include a listening condition to control for the effect of hearing the perturbation. It is well established that activation in STG is modulated by hearing sounds produced by others (Scott et al., 2004; 2009). Because perturbed feedback involves either having your own voice played back to you over headphones, or hearing additional masking sounds, it is possible that the response in the STG simply reflects the perception of these sounds, rather than speech production-specific processing.

The fMRI experiment described in chapter 5, investigating speech production in different masking sounds, addressed this issue by including a listening baseline in which participants heard the different masking sounds without vocalizing. The study attempted to address the question of whether the degree of mismatch between prediction and target modulates STG activity by using maskers that varied in their energetic potential (I.e. how effectively they occluded auditory feedback) as well as their informational content. This study found that, when the effect of hearing the different maskers was factored out, there were no significant differences in the STG between speaking in quiet and speaking in white noise- the strongest energetic masker, and therefore the condition that caused the greatest feedback perturbation. Rather, activity in STG was modulated in line with the informational content of the different masking sounds, and no areas that responded more to energetic masking were found in the univariate analysis. Taken as a whole, this evidence indicates that whilst the STG may have some role in processing and adapting to altered feedback in typical speakers, it is much more responsive to properties of our acoustic environment. The remaining two studies addressed the role of the STG in people with atypical voice control.

7.2. Do lesions involving the STG result in problems with feedback control?

If the STG is associated with feedback control, then damage to this region should result in impaired voice control. In chapter 3, this question was addressed with a case study of a 46-year-old man with expressive aphasia who reported impaired perception of his own voice following a left-sided stroke affecting posterior STG, as well as middle frontal cortex and insula. When speaking in white noise maskers at different intensity levels, he raised his vocal intensity, median pitch, and spectral centre of gravity consistently more than controls (indicating greater compensation for the noise), although this difference was not statistically significant. The percentage of unvoiced frames, used as a measure of vocal effort, did differ significantly between the patient and controls. The results were interpreted as demonstrating that the patient found noise more difficult to talk in and overcompensated for the perturbation, consistent with attenuated feedback perception. It is important to note that as the patient's lesion extended beyond temporal cortex into the insula, parietal and frontal cortex, it is not possible to make a definitive link between damage to the STG and feedback processing impairment. However, other studies (Singh & Schlanger, 1969; Boller et al., 1978) have found that people with similar lesions are more affected by perturbed feedback than controls, although they are less likely to correct errors in their speech at the semantic and phonological level. This dissociation suggests that there may be a difference between speech monitoring at a linguistic level (e.g. semantics and phonology) and at the acoustic level (e.g. intensity and speech timing).

7.3. DO PEOPLE WITH A HYPOTHESISED IMPAIRMENT IN FEEDBACK CONTROL DISPLAY ANOMALOUS STG ACTIVATION?

Stuttering is hypothesised to result from overreliance on auditory feedback (Civier et al., 2010). If this is so, then we might expect to see an absence of speaking induced suppression when people who stutter talk with normal feedback, compared to listening to voices. Conversely, when people who stutter talk in conditions that prevent them from relying on auditory feedback, we might expect fluency accompanied by comparative deactivation in the STG. In chapter 6, an fMRI study addressed this by looking at neural responses to speaking in two fluency enhancing conditions: synchronous speech and noise masking. Synchronous speech induces fluent speech reliably, whereas noise masking is only effective for some people who stammer. Compared to listening, there was no speaking-induced suppression response when participants spoke in quiet. In the white noise condition, activation in the STG was comparable to speaking in quiet and listening to voices. However, the STG showed a much greater response to speaking in synchrony with a partner than to speaking in the other conditions or to listening. An independent component analysis confirmed that activation in the STG was modulated significantly more by the synchronous speech conditions than for the other two speaking conditions, while activation in the basal ganglia (which have previously been implicated in stuttering) was modulated more strongly by speaking in quiet and speaking in noise masking than by the two synchrony conditions. Increased stuttering was also associated with activation in the basal ganglia. The lack of a speaking-induced suppression response when talking in quiet may indicate support for the theory that stammering results from an overreliance on auditory feedback, while the strong response in STG to synchronous

speech may reflect attention to the partner's voice, similar to the processing of unattended informational content found in chapter 5.

7.4. Discussion

We found some evidence that the STG is involved in speech monitoring and adaptation to perturbed feedback, particularly in people who have problems with voice control (Chapters 4 and 6). However, studies of typical speakers seem to suggest that this is not its primary role. Previous research has struggled to find a significant mismatch response to altered auditory feedback in the STG at whole brain level (Chapter 3). Meanwhile the two fMRI studies detailed in this thesis found that responses in the STG are significantly modulated by the amount of informational content in the background (Chapters 5 and 6). One issue with existing research on this topic is that different levels of speech monitoring and different types of hypothesised speech target are often conflated. That is, correction of semantic errors (such as word choice), phonemic errors (such as formant frequency) and acoustic or utterance-level errors (such as intensity) are all assumed to rely on the same monitoring process. However, there are many clear behavioural differences in the way that talkers deal with these different error types. Patients with expressive aphasia are less likely to correct semantic mistakes than controls, but over-compensate for feedback perturbations at the utterance level. Meanwhile, typical speakers rarely correct their semantic errors (Nooteboom, 1980), but reflexively raise their vocal intensity when it is attenuated by masking sounds (Lombard, 1911). Meanwhile, although talkers can adapt to shifts in formant frequency, compensation happens so slowly that it is necessary

for talkers to artificially prolong their utterances to demonstrate this response experimentally. It seems unlikely, therefore, that error monitoring is used in everyday speech to monitor our utterances at the phoneme or even the syllable level, as suggested in models of speech production (Guenther, 2006; Hickok, 2012). Rather, monitoring is a slow process which can correct for 'errors' at the utterance level but is unlikely to pick up missed targets at shorter latencies. Meanwhile, the STG is more likely to be processing unattended information than attending to the informational content of your own speech.

7.5. SUMMARY OF KEY FINDINGS:

In typical speakers:

- Contrasting perturbed with unperturbed feedback results in bilateral posterior STG activation
- However, the STG responds more to informational masking than to the quality of own-voice feedback.

In people with atypical vocal control:

- Lesions to the STG may result in difficulty processing auditory feedback at the utterance level.
- People who stutter lack a speaking-induced suppression response, potentially indicating an over-reliance on auditory feedback.
- However, fluency induced by synchronized speech is associated with a strong STG response; this may reflect an informational masking response.

Conclusions:

- The STG is involved in monitoring of the acoustic properties of speech at slow latencies.
- However, activation in the STG is more strongly modulated by informational properties of background sounds than by feedback quality.

Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *The Cerebellum*, *6*(3), 202–213. http://doi.org/10.1080/14734220701266742

Adams, M. R., Lewis, J. I., & Besozzi, T. E. (1973). The Effect of Reduced Reading Rate on Stuttering Frequency. *Journal of Speech Language and Hearing Research*, *16*(4), 671. http://doi.org/10.1044/jshr.1604.671

Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *NeuroImage*, *73*, 191–9. http://doi.org/10.1016/j.neuroimage.2012.08.020

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., … Levänen, S. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14608–14613. http://doi.org/10.1073/pnas.0510480103

Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, *1076*(1), 129–143. http://doi.org/10.1016/j.brainres.2005.11.104

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*. http://doi.org/10.1016/j.jcomdis.2004.03.001

Alm, P. A., & Risberg, J. (2007). Stuttering in adults: The acoustic startle response, temperamental traits, and biological factors. *Journal of Communication Disorders*, *40*(1), 1–41. http://doi.org/10.1016/j.jcomdis.2006.04.001

Altrows, I. F., & Bryden, M. P. (1977). Temporal factors in the effects of masking noise on fluency of stutterers. *Journal of Communication Disorders*, *10*(4), 315–329. http://doi.org/10.1016/0021-9924(77)90029-6

American Psychiatric Association. (2000). *DSM-IV*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR*. http://doi.org/10.1176

Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., Bauer, R. M., Nadeau, S., … Heilman, K. M. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain and Language*, *70*(1), 1–12. http://doi.org/10.1006/brln.1999.2135

Andrews, G., & Harris, M. (1964). *The syndrome of stuttering, Clinics in developmental medicine, No. 17.* London: William Heineman Medical Books Ltd.

Andrews, G., Howie, P. M., Dozsa, M., & Guitar, B. E. (1982). Stuttering: speech pattern characteristics under fluency-inducing conditions. *Journal of Speech and Hearing Research*, *25*(2), 208–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7120960

Armson, J., & Stuart, A. (1998). Effect of Extended Exposure to Frequency-Altered

Feedback on Stuttering During Reading and Monologue. *Journal of Speech Language and Hearing Research*, *41*(3), 479. http://doi.org/10.1044/jslhr.4103.479

Aubanel, V., & Cooke, M. (2013). Strategies adopted by talkers faced with fluctuating and competing speech maskers. *Journal of the Acoustic Society of America*.

Aubanel, V., Cooke, M., & Foster, E. (2013). Effects of the availability of visual information and presence of competing conversations on speech production. In *Interspeech* (pp. 4–7). Retrieved from http://91.203.57.208/upload/effects_of_the_availability_of_visual_information_and_presence_of_competing_conversations_on_speech_production.pdf

Backus, O. (1938). XLVIII Incidence of Stuttering among the Deaf. *Annals of Otology, Rhinology & Laryngology*, *47.3*, 632–635. Retrieved from http://aor.sagepub.com/content/47/3/632.full.pdf+html

Bakheit, A. M. O., Shaw, S., Carrington, S., & Griffiths, S. (2007). The rate and extent of improvement with therapy from the different types of aphasia in the first year after stroke. *Clinical Rehabilitation*, *21*(10), 941–9. http://doi.org/10.1177/0269215507078452

Barber, V. (1939). Studies in the Psychology of Stuttering, XV: Chorus reading as a distraction in stuttering. *Journal of Speech Disorders*, *4*(4), 371. http://doi.org/10.1044/jshd.0404.371

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bashford, J. A., Warren, R. M., & Brown, C. A. (1996). Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. *Perception & Psychophysics*, *58*(3), 342–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8935895

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, *109*, 418–428. http://doi.org/10.1016/j.neuroimage.2015.01.040

Belyk, M., Kraft, S. J., & Brown, S. (2015). Stuttering as a trait or state - an ALE meta-analysis of neuroimaging studies. *European Journal of Neuroscience*, *41*(2), 275–284. http://doi.org/10.1111/ejn.12765

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/486816

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300. http://doi.org/10.2307/2346101

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *The Journal of Speech Disorders*, *16*(1), 56–60.

Blakemore, S.-J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological Medicine*, *30*(5), 1131–1139. Retrieved from http://journals.cambridge.org/abstract_S0033291799002676

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–40. http://doi.org/10.1038/2870

Blesser, B. (1972). Speech Perception Under Conditions of Spectral Transformation: I. Phonetic Characteristics. *Journal of Speech Language and Hearing Research*, *15*(1), 5. http://doi.org/10.1044/jshr.1501.05

Blesser, B. A. (1969). Perception of spectrally rotated speech. Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/13625

Blomgren, M., Robb, M., & Chen, Y. (1998). A Note on Vowel Centralization in Stuttering and Nonstuttering Individuals. *Journal of Speech Language and Hearing Research*, *41*(5), 1042. http://doi.org/10.1044/jslhr.4105.1042

Bloodstein, O. (1948). A Rating Scale Study Of Conditions Under Which Stuttering Is Reduced Or Absent, 29–36.

Bloodstein, O. (1950). A Rating Scale Study Of Conditions Under Which Stuttering Is Reduced Or Absent. *Journal of Speech and Hearing Disorders*, *15*(1), 29. http://doi.org/10.1044/jshd.1501.29

Bloodstein, O. (2006). Some empirical observations about early stuttering: A possible link to language development. *Journal of Communication Disorders*, *39*(3), 185–191. http://doi.org/10.1016/j.jcomdis.2005.11.007

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language*, *9*(2), 153–170. http://doi.org/10.1016/0093-934X(80)90137-6

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer [software], *Available*.

Bogen, J. E., & Bogen, G. M. (1976). Wernicke's region- where is it? *Annals of the New York Academy of Sciences*, *280*(1 Origins and E), 834–843. http://doi.org/10.1111/j.1749-6632.1976.tb25546.x

Boller, F., Vrtunski, P. B., Kim, Y., & Mack, J. L. (1978). Delayed Auditory Feedback and Aphasia. *Cortex*, *14*(2), 212–226. http://doi.org/10.1016/S0010-9452(78)80047-1

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952

Braun, A., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., … Ludlow, C. L. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2(15)O positron emission tomography study. *Brain*, *120*(5), 761–784. http://doi.org/10.1093/brain/120.5.761

Brayton, E. R., & Conture, E. G. (1978). Effects of Noise and Rhythmic Stimulation on the Speech of Stutterers. *Journal of Speech Language and Hearing Research*, *21*(2), 285. http://doi.org/10.1044/jshr.2102.285

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

Brocklehurst, P. H. (2008). A Review of Evidence for the Covert Repair Hypothesis of Stuttering. *Contemporary Issues in Communication Science and Disorders*, *35*, 25–43.

Brown, A. S. (1991). A Review of the Tip-of-the-Tongue Experience. *Psychological Bulletin*, *109*(2), 204–223.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337. http://doi.org/10.1016/S0022-5371(66)80040-3

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 105–17. http://doi.org/10.1002/hbm.20140

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. http://doi.org/10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2001). Contralateral masking effects in dichotic listening with two competing talkers in the target ear. *The Journal of the Acoustical Society of America*, *109*(5), 2486–2486. http://doi.org/10.1121/1.4744845

Bryngelson, B. (1935). Sidedness as an Etiological Factor in Stuttering. *The Pedagogical Seminary and Journal of Genetic Psychology*, *47*(1), 204–217. http://doi.org/10.1080/08856559.1935.9943891

Buckingham, H. W. (2006). The Marc Dax (1770–1837)/Paul Broca (1824–1880) controversy over priority in science: Left hemisphere specificity for seat of articulate language and for lesions that cause aphemia. *Clinical Linguistics & Phonetics*, *20*(7–8), 613–619. http://doi.org/10.1080/02699200500266703

Budde, K. S., Barron, D. S., & Fox, P. T. (2014). *Stuttering, induced fluency, and natural fluency: A hierarchical series of activation likelihood estimation meta-analyses*. *Brain and Language* (Vol. 139).

Burke, B. D. (1969). Reduced auditory feedback and stuttering. *Behaviour Research and Therapy*, *7*(3), 303–308. http://doi.org/10.1016/0005-7967(69)90011-4

Burke, B. D. (1975). Susceptibility to delayed auditory feedback and dependence on auditory or oral sensory feedback. *Journal of Communication Disorders*, *8*(1), 75–96. http://doi.org/10.1016/0021-9924(75)90028-3

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9637026

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's Aphasia Is Not Associated with a Single Pattern of Comprehension Performance. *Brain and Language*, *76*(2), 158–184. http://doi.org/10.1006/brln.1999.2275

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the `tip-of-the-tongue' phenomenon. *Cognition*, *64*(3), 309–343. http://doi.org/10.1016/S0010-0277(97)00031-0

Carhart, R. (1969). Perceptual Masking in Multiple Sound Backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. http://doi.org/10.1121/1.1911445

Chang, S.-E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage*, *39*(3), 1333–1344. http://doi.org/10.1016/j.neuroimage.2007.09.067

Chang, S.-E., & Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain : A Journal of Neurology*, *136*(Pt 12), 3709–26. http://doi.org/10.1093/brain/awt275

Cherry, C., & Sayers, B. M. (1956). Experiments upon the total inhibition of stammering by external control, and some clinical results. *Journal of Psychosomatic Research*, *1*(4), 233–246. http://doi.org/10.1016/0022-3999(56)90001-0

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. http://doi.org/10.1121/1.1907229

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human Brain Mapping*, *28*(9), 868–79. http://doi.org/10.1002/hbm.20315

Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PloS One*, *6*(5), e18307. http://doi.org/10.1371/journal.pone.0018307

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, *35*(3), 246–279. http://doi.org/10.1016/j.jfludis.2010.05.002

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics

in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. http://doi.org/10.1016/S0022-5371(73)80014-3

Connally, E. L., Ward, D., Howell, P., & Watkins, K. E. (2014). Disrupted white matter in language and motor tracts in developmental stuttering. *Brain and Language*, *131*, 25–35. http://doi.org/10.1016/j.bandl.2013.05.013

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. http://doi.org/10.1121/1.2166600

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059–69. http://doi.org/10.1121/1.3478775

Costello Ingham, J. (1983). Current status of stuttering and behavior modification-I: Recent trends in the application of behavior modification in children and adults. *Journal of Fluency Disorders*, *18*(1), 27–55. http://doi.org/10.1016/0094-730X(83)90004-9

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*, 1196–1208.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245–260. http://doi.org/10.1080/02643294.2010.513967

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. *Experimental Brain Research*. Retrieved from http://link.springer.com/article/10.1007/BF00249600

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, *31*(2), 139–148. http://doi.org/10.1016/S0095-4470(02)00082-7

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*(1), 16–28. http://doi.org/10.1016/j.wocn.2008.08.003

Cykowski, M. D., Kochunov, P. V., Ingham, R. J., Ingham, J. C., Mangin, J.-F., Riviere, D., … Fox, P. T. (2008). Perisylvian Sulcal Morphology and Cerebral Asymmetry Patterns in Adults Who Stutter. *Cerebral Cortex*, *18*(3), 571–583. http://doi.org/10.1093/cercor/bhm093

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. http://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629.

http://doi.org/10.1016/S0022-5371(81)90202-4

Dreher, J. J., & O'Neill, J. (1957). Effects of Ambient Noise on Speaker Intelligibility for Words and Phrases. *The Journal of the Acoustical Society of America*, *29*(12), 1320. http://doi.org/10.1121/1.1908780

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159–161. http://doi.org/10.1038/384159a0

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain*. http://doi.org/10.1093/brain/awm042

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1–2), 145–177. http://doi.org/10.1016/j.cognition.2003.11.002

Duffy, J. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences. Retrieved from https://books.google.com/books?hl=en&lr=&id=ATARAAAAQBAJ&oi=fnd&pg=PP1&dq=duffy+1995+motor+speech+disorders&ots=LwfC-x7rMl&sig=g2iLj9v9eD1vDqmMtI2BFFwCRPw

Dworzynski, K., Remington, A., Rijsdijk, F., Howell, P., Plomin, R., G., A. N., … R., T. (2007). Genetic Etiology in Cases of Recovered and Persistent Stuttering in an Unselected, Longitudinal Sample of Young Twins. *American Journal of Speech-Language Pathology*, *16*(2), 169. http://doi.org/10.1044/1058-0360(2007/021)

Egnor, S. E. R., Wickelgren, J. G., & Hauser, M. D. (2007). Tracking silence: adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, *193*(4), 477–83. http://doi.org/10.1007/s00359-006-0205-7

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., … Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70–85. http://doi.org/10.1016/j.neuroimage.2016.04.072

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7900–5. http://doi.org/10.1073/pnas.1602413113

Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, *89*(4), 2194–207. http://doi.org/10.1152/jn.00627.2002

Eliades, S. J., & Wang, X. (2012). Neural correlates of the lombard effect in primate auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for*

Neuroscience, *32*(31), 10737–48. http://doi.org/10.1523/JNEUROSCI.3448-11.2012

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, *28*(3), 483–500. http://doi.org/10.1162/jocn_a_00913

Fairbanks, G. (1954). Systematic Research In Experimental Phonetics:* 1. A Theory Of The Speech Mechanism As A Servosystem. *Journal of Speech and Hearing Disorders*, *19*(2), 133. http://doi.org/10.1044/jshd.1902.133

Fairbanks, G. (1960). The rainbow passage. In *Voice and articulation drillbook, 2.*

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech- reception threshold for impaired and normal hearing, *88*(4), 1725–1736.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(49), 16643–50. http://doi.org/10.1523/JNEUROSCI.1809-10.2010

Foundas, A. L., Bollich, A. M., Corey, D. M., Hurley, M., & Heilman, K. M. (2001). Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. *Neurology*, *57*(2), 207–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11468304

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C., & Heilman, K. M. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. *Neurology*, *63*(9), 1640–1646. http://doi.org/10.1212/01.WNL.0000142993.33158.2A

Foundas, A. L., Corey, D. M., Angeles, V., Bollich, A. M., Crabtree-Hartman, E., & Heilman, K. M. (2003). Atypical cerebral laterality in adults with persistent developmental stuttering. *Neurology*, *61*(10), 1378–85. http://doi.org/10.1212/01.WNL.0000094320.44334.86

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–8. http://doi.org/10.1073/pnas.0504136102

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., … Lancaster, J. L. (1996). A PET study of the neural systems of stuttering. *Nature*, *382*(6587), 158–162. http://doi.org/10.1038/382158a0

Franklin, G. F., Powell, J. D., & Emami-Naeini, A. (2002). *Feedback Control of Dynamic Systems*. *Sound And Vibration* (Vol. 7). Retrieved from http://www.pearsonhighered.com/educator/product/Feedback-Control-of-Dynamic-Systems-6E/9780136019695.page

Freud, S. (1915). Psychopathology of Everyday Life. *The Journal of Nervous and Mental Disease*, *42*(9), 655. http://doi.org/10.1097/00005053-191509000-00012

Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2015). Chronic Broca's Aphasia Is Caused by Damage to Broca's and Wernicke's Areas. *Cerebral Cortex (New York, N.Y. : 1991)*, *25*(12), 4689–96. http://doi.org/10.1093/cercor/bhu152

Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *The British Journal of Psychiatry*, *153*(4), 437–443. http://doi.org/10.1192/bjp.153.4.437

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances, *47*(1), 27–52. Retrieved from http://links.jstor.org/sici?sici=0097-8507%28197103%2947%3A1%3C27%3ATNNOAU%3E2.0.CO%3B2-M

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., … McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977. http://doi.org/10.1093/cercor/bhj039

Galaburda, A., LeMay, M., Kemper, T., & Geschwind, N. (1978). Right-left asymmetrics in the brain. *Science*, *199*(4331).

Garber, S. F., & Martin, R. R. (1974). The Effects of White Noise on the Frequency of Stuttering. *Journal of Speech Language and Hearing Research*, *17*(1), 73. http://doi.org/10.1044/jshr.1701.73

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing …*, *53*(June), 588–608. Retrieved from http://jslhr.pubs.asha.org/article.aspx?articleid=1781559

Garrett, M. (1992). Lexical retrieval processes: Semantic field effects. *Frames, Fields and Contrasts: New Essays in*. Retrieved from https://books.google.com/books?hl=en&lr=&id=ZGrEEnE7mI4C&oi=fnd&pg=PA377&dq=garrett+1992+word+substitutions&ots=3zGVtlhvoq&sig=qGgxhgmgkaZF5PiocEF5XrsCMZM

Geschwind, N., & Galaburda, A. M. (1985). Cerebral Lateralization. *Archives of Neurology*, *42*(5), 428. http://doi.org/10.1001/archneur.1985.04060050026008

Giraud, A.-L., Neumann, K., Bachoud-Levi, A.-C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language*, *104*(2), 190–199. http://doi.org/10.1016/j.bandl.2007.04.005

Gray, H. (1918). *Anatomy of the human body*. *Bartleby.Com, May 2000*. Retrieved from www.bartleby.com/107/

Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M. F., … Smith, S. M. (2016). Hand classification of fMRI ICA noise components.

NeuroImage. http://doi.org/10.1016/j.neuroimage.2016.12.036

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*. http://doi.org/10.1016/S0166-2236(02)02191-4

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–65. http://doi.org/10.1016/j.jcomdis.2006.06.013

Guenther, F. H., & Hickok, G. (2015). Chapter 9 – Role of the auditory system in speech production. In *Handbook of Clinical Neurology* (Vol. 129, pp. 161–175). http://doi.org/10.1016/B978-0-444-62630-1.00009-3

Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., Jones, K., O., B., … L., W. R. (1994). Effect of Frequency-Altered Feedback on Stuttering Frequency at Normal and Fast Speech Rates. *Journal of Speech Language and Hearing Research*, *37*(6), 1313. http://doi.org/10.1044/jshr.3706.1313

Harms, M. A., & Malone, J. Y. (1939). The Relationship of Hearing Acuity to Stammering. *Journal of Speech Disorders*, *4*(4), 363. http://doi.org/10.1044/jshd.0404.363

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Human Brain Mapping*, *20*(1), 22–8. http://doi.org/10.1002/hbm.10119

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience*, *13*(2), 135–45. http://doi.org/10.1038/nrn3158

Hickok, G. (2014a). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, *29*(1), 2–20. http://doi.org/10.1080/01690965.2013.834370

Hickok, G. (2014b). Toward an Integrated Psycholinguistic, Neurolinguistic, Sensorimotor Framework for Speech Production. *Language and Cognitive Processes*, *29*(1), 52–59. http://doi.org/10.1080/01690965.2013.852907

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. http://doi.org/10.1162/089892903322307393

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402. http://doi.org/10.1038/nrn2113

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., … Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport*, *8*(9–10), 2379–2382. http://doi.org/10.1097/00001756-199707070-00055

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the

Auditory Cortex during Speech: An MEG Study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138. http://doi.org/10.1162/089892902760807140

Howell, P. (1990). Changes in Voice Level Caused by Several Forms of Altered Feedback in Fluent Speakers and Stutterers. *Language Ans Speech*, *33*(4), 325–338. http://doi.org/10.1177/002383099003300402

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. *IN THE THEORY AND HISTORY OF …*. Retrieved from https://books.google.com/books?hl=en&lr=&id=4YE5AAAAQBAJ&oi=fnd&pg=PA75&dq=howell+explan+stuttering&ots=W5aZgNc58-&sig=vDvUnJbbgQ1UQj8uXk16Un8d5cA

Howell, P., El-Yaniv, N., & Powell, D. J. (1987). Factors Affecting Fluency in Stutterers when Speaking under Altered Auditory Feedback. In *Speech Motor Dynamics in Stuttering* (pp. 361–369). Vienna: Springer Vienna. http://doi.org/10.1007/978-3-7091-6969-8_28

Howell, P., & Powell, D. J. (1987). Delayed auditory feedback with delayed sounds varying in duration. *Perception & Psychophysics*, *42*(2), 166–172. http://doi.org/10.3758/BF03210505

Howell, P., & Sackin, S. (2000). Speech Rate Modification and Its Effects on Fluency Reversal in Fluent Speakers and People Who Stutter. *Journal of Developmental and Physical Disabilities*, *12*(4), 291–315. http://doi.org/10.1023/A:1009428029167

Hurford, J. R. (2004). Human uniqueness, learned symbols and recursive thought. *European Review*, *12*(4), 551–565. Retrieved from http://ideas.repec.org/a/cup/eurrev/v12y2004i04p551-565_00.html

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. In *The new cognitive neurosciences* (pp. 845–865).

Jäncke, L., Hänggi, J., & Steinmetz, H. (2004). Morphological brain differences between adult stutterers and non-stutterers. *BMC Neurology*, *4*(1), 23. http://doi.org/10.1186/1471-2377-4-23

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and Joint Speech: Right Hemisphere Contributions to Synchronized Vocal Production. *Journal of Neuroscience*, *36*(17), 4669–4680. http://doi.org/10.1523/JNEUROSCI.4075-15.2016

Jesteadt, W., Bacon, S. P., Lehman, J. R., Jesteadt, W., & Baconf, S. P. (1982). Forward masking as a function of frequency, masker level, and signal delay. *The Journal of the Acoustical Society of America*, *71*(950). http://doi.org/10.1121/1.387576

Johnson, W., & Rosen, L. (1937). Studies in the psychology of stuttering: Effect of certain changes in speech pattern upon frequency of stuttering. *Journal of Speech and Hearing Research*, *2*, 105–109.

Junqua, J. (1993). The Lombard reflex and its role on human and automatic speech recognizers, 510–524.

Kalinowski, J., Armson, J., Stuart, R.-M. A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *LANGUAGE AND SPEECH*, *36*(1), 1–16.

Kalinowski, J., & Saltuklaroglu, T. (2003). Speaking with a mirror: engagement of mirror neurons via choral speech and its derivatives induces stuttering inhibition. *Medical Hypotheses*, *60*(4), 538–543. http://doi.org/10.1016/S0306-9877(03)00004-5

Kalinowski, J., Stuart, A., Sark, S., & Armson, J. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. *International Journal of Language & Communication Disorders*, *31*(3), 259–269. http://doi.org/10.3109/13682829609033157

Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H., & Giraud, A.-L. (2009). How the brain repairs stuttering. *Brain*, *132*(10), 2747–2760. http://doi.org/10.1093/brain/awp185

Kertesz, A., & McCabe, P. (1977). Recovery patterns and prognosis in aphasia. *Brain*, *100*, 1–18.

Kiefte, M., & Armson, J. (2008). Dissecting choral speech: Properties of the accompanist critical to stuttering reduction. *Journal of Communication Disorders*, *41*(1), 33–48. http://doi.org/10.1016/j.jcomdis.2007.03.002

King, H. (2012). Antiphon: Notes on the People's Microphone*. *Journal of Popular Music Studies*, *24*(2), 238–246. http://doi.org/10.1111/j.1533-1598.2012.01327.x

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., … Senda, M. (2005). Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner. *Acoustical Science and Technology*, *26*(5), 465–468. http://doi.org/10.1250/ast.26.465

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. http://doi.org/10.1038/nn.2303

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(27), 9351–8. http://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane, H., & Tranel, B. (1971). The Lombard Sign and the role of Hearing in Speech. *Journal of Speech and Hearing Research*, *14*, 677–710.

Lane, H., Tranel, B., & Sisson, C. (1970). Regulation of Voice Communication by Sensory Dynamics. *Journal of the Acoustic Society of America*, *47*(2), 618–624.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to Intention. *Science*, *303*(5661), 1208–1210. http://doi.org/10.1126/science.1090973

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. http://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: Bradford. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0,5&cluster=84816083015255 7898

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. http://doi.org/10.1016/S1364-6613(99)01319-4

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. http://doi.org/10.1017/S0140525X99001776

Lichtheim, L. (1885). On aphasia. *Brain*, *1*(6), 1347–1350. http://doi.org/10.1093/brain/awl134

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends in Neurosciences*, *39*(8), 527–542. http://doi.org/10.1016/j.tins.2016.06.003

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Auditory feedback of one's own voice is used for high-level semantic monitoring: the "self-comprehension" hypothesis. *Frontiers in Human Neuroscience*, *8*(March), 166. http://doi.org/10.3389/fnhum.2014.00166

Lombard, E. (1911). Le signe de l'elevation de la voix. *Annales Des Maladies de L'Oreille et Du Larynx*, *37*, 101–119.

Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., … Lin, C. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. *Cortex*, *46*(1), 49–67. http://doi.org/10.1016/j.cortex.2009.02.017

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, *124*(5), 3261–75. http://doi.org/10.1121/1.2990705

Lund, T. E., Nørgaard, M. D., Rostrup, E., Rowe, J. B., & Paulson, O. B. (2005). Motion or activity: their role in intra- and inter-subject variation in fMRI. *NeuroImage*, *26*(3), 960–4. http://doi.org/10.1016/j.neuroimage.2005.02.021

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*(1), 29–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2317599

Macleod, J., Kalinowski, J., Stuart, A., & Armson, J. (1995). Effect of single and combined altered auditory feedback on stuttering frequency at two speech rates. *Journal of Communication Disorders*, *28*(3), 217–228. http://doi.org/10.1016/0021-9924(94)00010-W

Maher, L. M., Rothi, L. J. G., & Heilman, K. M. (1994). Lack of Error Awareness in an Aphasic Patient with Relatively Preserved Auditory Comprehension. *Brain and Language*, *46*(3), 402–418. http://doi.org/10.1006/brln.1994.1022

Månsson, H. (2000). Childhood stuttering: Incidence and development. *Journal of Fluency Disorders*, *25*(1), 47–57. http://doi.org/10.1016/S0094-730X(99)00023-6

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. http://doi.org/10.1016/j.cogpsych.2009.04.001

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Internal Models and Feedback Bias in Stuttering E Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering. *Contemporary Issues in Communication Disorders*, *31*, 105–122.

Max, L., & Yudman, E. M. (2003). Accuracy and Variability of Isochronous Rhythmic Timing Across Motor Systems in Stuttering Versus Nonstuttering Individuals. *Journal of Speech Language and Hearing Research*, *46*(1), 146. http://doi.org/10.1044/1092-4388(2003/012)

McGuire, P. K., Silbersweig, D. a., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *European Psychiatry*, *11*, 182s–183s. http://doi.org/10.1016/0924-9338(96)88510-5

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S. K. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *The Journal of the Acoustical Society of America*, *140*(1), 8–19. http://doi.org/10.1121/1.4948587

Milisen, R., & Johnson, W. (1936). A comparative study of stutterers, former stutterers, and normal speakers whose handedness has been changed. *Archives of Speech*, *1*, 61–68.

Mink, J. W. (2003). The Basal Ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Archives of Neurology*, *60*(10), 1365–8. http://doi.org/10.1001/archneur.60.10.1365

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, T. P., Oxman, A., … Hopewell, S. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. http://doi.org/10.1371/journal.pmed.1000097

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R.

(1978). Broca aphasia: pathologic and clinical. *Neurology*, *28*(4), 311–24. http://doi.org/10.1212/WNL.28.4.311

Morosan, P., Schleicher, A., Amunts, K., & Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. In *Anatomy and Embryology* (Vol. 210, pp. 401–406). http://doi.org/10.1007/s00429-005-0029-1

Murphy, K., Corfield, D. R., Guz, A., Fink, G. R., Wise, R. J. S., Harrison, J., & Adams, L. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, *83*(5).

Murray, F. P. (1969). An investigation of variably induced white noise upon moments of stuttering. *Journal of Communication Disorders*, *2*(2), 109–114. http://doi.org/10.1016/0021-9924(69)90034-3

Murry, T. (1990). Pitch-matching accuracy in singers and nonsingers. *Journal of Voice*, *4*(4), 317–321. http://doi.org/10.1016/S0892-1997(05)80048-7

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*(12), 1362–1368. http://doi.org/10.1093/cercor/bhg083

Natke, U. (2000). Reduction of stuttering frequency using frequency-shifted and delayed auditory feedback. *Folia Phoniatrica et Logopaedica : Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *52*(4), 151–9. http://doi.org/21530

Neef, N. E., Anwander, A., & Friederici, A. D. (2015). The Neurobiological Grounding of Persistent Stuttering: from Structure to Function. *Current Neurology and Neuroscience Reports*, *15*(9), 63. http://doi.org/10.1007/s11910-015-0579-4

Neef, N. E., Hoang, T. N. L., Neef, A., Paulus, W., & Sommer, M. (2015). Speech dynamics are coded in the left motor cortex in fluent speakers but not in adults who stutter. *Brain*, *138*(3), 712–725. http://doi.org/10.1093/brain/awu390

Neelley, J. N. (1961). A study of the speech behavior of stutterers and nonstutterers under normal and delayed auditory feedback. *The Journal of Speech and Hearing Disorders*, *(Suppl 7)*, 63–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13728119

Neumann, K., Euler, H. A., Gudenberg, A. W. von, Giraud, A.-L., Lanfermann, H., Gall, V., & Preibisch, C. (2003). The nature and treatment of stuttering as revealed by fMRI: A within- and between-group comparison. *Journal of Fluency Disorders*, *28*(4), 381–410. http://doi.org/10.1016/j.jfludis.2003.07.003

Neumann, K., Preibisch, C., Euler, H. A., Gudenberg, A. W. von, Lanfermann, H., Gall, V., & Giraud, A.-L. (2005). Cortical plasticity associated with stuttering therapy. *Journal of Fluency Disorders*, *30*(1), 23–39. http://doi.org/10.1016/j.jfludis.2004.12.002

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5),

419–446. http://doi.org/10.1191/0962280203sm341ra

Nippold, M. A. (2002). Stuttering and Phonology. *American Journal of Speech-Language Pathology*, *11*(2), 99. http://doi.org/10.1044/1058-0360(2002/011)

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, *33*(29), 12090–8. http://doi.org/10.1523/JNEUROSCI.1008-13.2013

Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., & Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research*, *29*(4), 283–289. http://doi.org/10.1016/S0168-0102(97)00097-7

Nooteboom, S. G. (1980). Speaking and unspeaking : detection and correction of phonological and lexical errors in spontaneous speech. In V. Fromkin (Ed.), *Errors in linguistic performance : slips of the tongue, ear, pen, and hand* (pp. 87–95). Academic Press.

Nota, Y., & Honda, K. (2004). Brain regions involved in motor control of speech. *Acoustical Science and Technology*, *25*(4), 286–289. http://doi.org/10.1250/ast.25.286

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, *17*(10), 2251–2257. http://doi.org/10.1093/cercor/bhl133

Oomen, C. C., Postma, A., & Kolk, H. H. (2001). Prearticulatory and Postarticulatory Self-Monitoring in Broca's Aphasia. *Cortex*, *37*(5), 627–641. http://doi.org/10.1016/S0010-9452(08)70610-5

Osser, H., & Peng, F. (1964a). A Cross Cultural Study of Speech Rate. *Language and Speech*, *7*(2), 120–125. http://doi.org/10.1177/002383096400700208

Osser, H., & Peng, F. (1964b). A Cross Cultural Study of Speech Rate. *Language and Speech*, *7*(2), 120–125. http://doi.org/10.1177/002383096400700208

Pai, M.-C. (1999). *Supplementary motor area aphasia: a case report*. *Clinical Neurology and Neurosurgery* (Vol. 101).

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. http://doi.org/10.1016/j.neuroimage.2012.02.068

Patel, R., Schell, K. W., A., C., C., J. J., C., J. J., E., L., … L., W. A. (2008). The Influence of Linguistic Content on the Lombard Effect. *Journal of Speech Language and Hearing Research*, *51*(1), 209. http://doi.org/10.1044/1092-4388(2008/016)

Pedersen, P. M., Vinter, K., & Olsen, T. S. (2004). Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovascular Diseases (Basel, Switzerland)*, *17*(1), 35–43. http://doi.org/10.1159/000073896

Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. *Frontiers in Human Neuroscience*, *6*, 309. http://doi.org/10.3389/fnhum.2012.00309

Perkins, W. H., Kent, R. D., & Curlee, R. F. (1991). A Theory of Neuropsycholinguistic Function in Stuttering. *Journal of Speech Language and Hearing Research*, *34*(4), 734. http://doi.org/10.1044/jshr.3404.734

Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology*, *11*(6), 663–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11741015

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3795886

Pick, H. L., Siegel, G. M., Fox, P. W., & Kearney, J. K. (1989). Inhibiting the Lombard effect, 894–900.

Pickett, E. R., Kuniholm, E., Protopapas, A., Friedman, J., & Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: a case study. *Neuropsychologia*, *36*(2), 173–188. http://doi.org/10.1016/S0028-3932(97)00065-1

Pilgrim, L. K., Fadili, J., Fletcher, P., & Tyler, L. K. (2002). Overcoming Confounds of Stimulus Blocking: An Event-Related fMRI Design of Semantic Processing. *NeuroImage*, *16*(3), 713–723. http://doi.org/10.1006/nimg.2002.1105

Pittman, A. L., & Wiley, T. L. (2001). Recognition of Speech Produced in Noise. *Journal of Speech, Language, & Hearing Research*, *44*(3), 487–496. http://doi.org/10.1044/1092-4388(2001/038)

Plant, R. L., & Younger, R. M. (2000). The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. *Journal of Voice*, *14*(2), 170–177. http://doi.org/10.1016/S0892-1997(00)80024-7

Pollack, I., & Pickett, J. M. (1957). Cocktail Party Effect. *The Journal of the Acoustical Society of America*, *29*(11), 1262–1262. http://doi.org/10.1121/1.1919140

Porter, G., & Howard, D. (2004). CAT: comprehensive aphasia test. Psychology Press Ltd.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, *36*(3), 472–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8331905

Price, C. J., & Crinion, J. (2005). The latest on functional imaging studies of aphasic stroke. *Current Opinion in Neurology*, *18*(4), 429–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16003120

Proctor, A., Yairi, E., Duff, M. C., Zhang, J., Association, A. S.-L.-H., B., A., … C., C. (2008). Prevalence of Stuttering in African American Preschoolers. *Journal of Speech Language and Hearing Research*, *51*(6), 1465. http://doi.org/10.1044/1092-4388(2008/07-0057)

Rafii, Z., & Pardo, B. (2011). A simple music/voice separation method based on the extraction of the repeating musical structure. In *36th International Conference on Acoustics, Speech and Signal Processing* (pp. 1–4). Retrieved from http://music.cs.northwestern.edu/publications/Rafii-Pardo - A Simple Music-Voice Separation Method based on the Extraction of the Repeating Musical Structure - ICASSP 2011.pdf

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–24. http://doi.org/10.1038/nn.2331

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, *64*(4), 700–706. http://doi.org/10.1212/01.WNL.0000152156.90779.89

Riley, G. D. (1972). A Stuttering Severity Instrument for Children and Adults. *Journal of Speech and Hearing Disorders*, *37*(3), 314. http://doi.org/10.1044/jshd.3703.314

Robb, M., & Blomgren, M. (1997). Analysis of F2 transitions in the speech of stutterers and nonstutterers. *Journal of Fluency Disorders*, *22*(1), 1–16. http://doi.org/10.1016/S0094-730X(96)00016-2

Rosenfield, D. B. (1980). Cerebral dominance and stuttering. *Journal of Fluency Disorders*, *5*(3), 171–185. http://doi.org/10.1016/0094-730X(80)90027-3

Salmelin, R., Schnitzler, A., Schmitz, F., Jäncke, L., Witte, O. W., & Freund, H. J. (1998). Functional organization of the auditory cortex is different in stutterers and fluent speakers. *Neuroreport*, *9*(10), 2225–9. http://doi.org/10.1097/00001756-199807130-00014

Salomon, G., & Starr, A. (1963). ELECTROMYOGRAPHY OF MIDDLE EAR MUSCLES IN MAN DURING MOTOR ACTIVITIES. *Acta Neurologica Scandinavica*, *39*(2), 161–168. http://doi.org/10.1111/j.1600-0404.1963.tb05317.x

Sato, Y., Mori, K., Koizumi, T., Minagawa-Kawai, Y., Tanaka, A., Ozawa, E., … Mazuka, R. (2011). Functional Lateralization of Speech Processing in Adults and Children Who Stutter. *Frontiers in Psychology*, *2*, 70. http://doi.org/10.3389/fpsyg.2011.00070

Schievink, W. I., Limburg, M., Oorthuys, J. W., Fleury, P., & Pope, F. M. (1990). Cerebrovascular disease in Ehlers-Danlos syndrome type IV. *Stroke*, *21*(4).

Schlenck, K. J., Huber, W., & Willmes, K. (1987). “Prepairs” and repairs: different monitoring functions in aphasic language production. *Brain and Language*, *30*(2),

226–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2436704

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain*, *132*(12), 3411–3427. http://doi.org/10.1093/brain/awp284

Scott, S. K. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. http://doi.org/10.1093/brain/123.12.2400

Scott, S. K. (2012). The neurobiology of speech perception and production--can functional imaging tell us anything we did not already know? *Journal of Communication Disorders*, *45*(6), 419–25. http://doi.org/10.1016/j.jcomdis.2012.06.007

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligable speech in the left temporal lobe. *Brain*, *123*, 2400–2406.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., & Wise, R. J. S. (2009). The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *The Journal of the Acoustical Society of America*, *125*(3), 1737–43. http://doi.org/10.1121/1.3050255

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, *115*(2), 813. http://doi.org/10.1121/1.1639336

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, *115*(2), 813–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15000192

Shahed, J., & Jankovic, J. (2001). Re-emergence of childhood stuttering in Parkinson's disease: a hypothesis. *Movement Disorders : Official Journal of the Movement Disorder Society*, *16*(1), 114–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11215569

Shergill, S. S., Bullmore, E., Simmons, A., Murray, R., & McGuire, P. (2014). Functional Anatomy of Auditory Verbal Imagery in Schizophrenic Patients With Auditory Hallucinations. Retrieved from http://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.157.10.1691

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. http://doi.org/10.1016/j.tics.2008.02.003.Object-based

Singh, S., & Schlanger, B. B. (1969). Effects of Delayed Sidetone On the Speech of Aphasic, Dysarthric, and Mentally Retarded Subjects. *Language and Speech*, *12*(3), 167–174. http://doi.org/10.1177/002383096901200303

Smith, L., & Klein, R. (1990). Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 852. http://doi.org/10.1037/0278-7393.16.5.852

Sommer, M., Knappmeyer, K., Hunter, E. J., Gudenberg, A. W., Neef, N., & Paulus, W. (2009). Normal interhemispheric inhibition in persistent developmental stuttering. *Movement Disorders*, *24*(5), 769–773. http://doi.org/10.1002/mds.22383

Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Büchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. *The Lancet*, *360*(9330), 380–383. http://doi.org/10.1016/S0140-6736(02)09610-1

Sparks, G., Grant, D. E., Millay, K., Walker-Batson, D., & Hynan, L. S. (2002). The effect of fast speech rate on stuttering frequency during delayed auditory feedback. *Journal of Fluency Disorders*, *27*(3), 187–201. http://doi.org/10.1016/S0094-730X(02)00128-6

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition*, *35*(2), 123–157. http://doi.org/10.1016/0010-0277(90)90012-9

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., & Moore, B. C. J. (2011). The importance for speech intelligibility of random fluctuations in "steady" background noise. *The Journal of the Acoustical Society of America*, *130*(5), 2874–81. http://doi.org/10.1121/1.3641371

Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317–26. http://doi.org/10.1121/1.4725766

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, *111*(5), 2237. http://doi.org/10.1121/1.1466868

Summers, W. Van, Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustic Society of America*, *84*(3), 917–928. http://doi.org/10.1016/j.amjcard.2008.06.048.Long-term

Sutton, S., & Chase, R. A. (1961). White noise and stuttering. *Journal of Speech & Hearing Research*, *4*, 72.

Takaso, H., Eisner, F., Wise, R. J. S., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *Journal of Speech, Language, and Hearing Research : JSLHR*, *53*(2), 226–36. http://doi.org/10.1044/1092-4388(2009/09-0009)

Tourville, J. a, Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–43. http://doi.org/10.1016/j.neuroimage.2007.09.054

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *NeuroImage*, *57*(4), 1507–16. http://doi.org/10.1016/j.neuroimage.2011.05.039

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*(2), 499–503. http://doi.org/10.1016/j.neuroscience.2007.02.023

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, *35*(10), 1319–1327. http://doi.org/10.1016/S0028-3932(97)00085-7

Travis, L. E. (1978). The cerebral dominance theory of stuttering: 1931--1978. *The Journal of Speech and Hearing Disorders*, *43*(3), 278–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/357839

Trupe, L. A., Varma, D. D., Gomez, Y., Race, D., Leigh, R., Hillis, A. E., & Gottesman, R. F. (2013). Chronic Apraxia of Speech and Broca's Area. *Stroke*, *44*(3), 740–744. http://doi.org/10.1161/STROKEAHA.112.678508

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. *Human Brain Mapping*, *33*(1), 1–13. http://doi.org/10.1002/hbm.21186

Upadhyay, J., Silver, A., Knaus, T. a, Lindgren, K. a, Ducros, M., Kim, D.-S., & Tager-Flusberg, H. (2008). Effective and structural connectivity in the human auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(13), 3341–3349. http://doi.org/10.1523/JNEUROSCI.4434-07.2008

van de Ven, V., Esposito, F., & Christoffels, I. K. (2009). Neural network of speech monitoring overlaps with overt speech production and comprehension networks: a sequential spatial and temporal ICA study. *NeuroImage*, *47*(4), 1982–91. http://doi.org/10.1016/j.neuroimage.2009.05.057

Van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., & Clarke, S. (2011). Where sound position influences sound object representations: A 7-T fMRI study. *NeuroImage*, *54*(3), 1803–1811. http://doi.org/10.1016/j.neuroimage.2010.10.032

Varadarajan, V. S., & Hansen, J. H. L. (2006). Analysis of Lombard effect under different types and levels of noise with application to in-set speaker ID systems. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 2, pp. 937–940).

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical Gender Is on the Tip of Italian Tongues. *Source: Psychological Science*, *8*(4), 314–317. http://doi.org/10.1111/j.1467-9280.1997.tb00444.x

Vrtunski, P. B., Mack, J. L., Boller, F., & Kim, Y. (1976). Response to Delayed Auditory

Feedback in Patients with Hemispheric Lesions. *Cortex*, *12*(4), 395–404. http://doi.org/10.1016/S0010-9452(76)80043-3

Warhurst, S., Madill, C., McCabe, P., Heard, R., & Yiu, E. (2012). The vocal clarity of female speech-language pathology students: an exploratory study. *Journal of Voice : Official Journal of the Voice Foundation*, *26*(1), 63–8. http://doi.org/10.1016/j.jvoice.2010.10.008

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences*, *28*(12), 636–643. http://doi.org/10.1016/j.tins.2005.09.010

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2007). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, *131*(1).

Webster, J. C., & Klumpp, R. G. (1962). Effects of ambient noise and nearby talkers on a face-to-face communication task. *Journal of the Acoustic Society of America*, *34*(7), 936–941.

Weinstein, E. A., Lyerly, O. G., Cole, M., & Ozer, M. N. (1966). Meaning in Jargon Aphasia. *Cortex*, *2*(2), 165–187. http://doi.org/10.1016/S0010-9452(66)80001-1

Wernicke, C. (1874). *Der Aphasische Symptomencomplex*. *Breslau. Max Cohn & Weigert*. http://doi.org/10.1007/978-3-642-65950-8

Wingate, M. E. (1964). A Standard Definition of Stuttering. *Journal of Speech and Hearing Disorders*, *29*(4), 484. http://doi.org/10.1044/jshd.2904.484

Wingate, M. E. (1970). Effect on Stuttering of Changes in Audition. *Journal of Speech Language and Hearing Research*, *13*(4), 861. http://doi.org/10.1044/jshr.1304.861

Wise, R. J., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, *353*(9158), 1057–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10199354

Wood, G., Nuerk, H.-C., Sturm, D., & Willmes, K. (2008). Using parametric regressors to disentangle properties of multi-feature processes. *Behavioral and Brain Functions : BBF*, *4*(1), 38. http://doi.org/10.1186/1744-9081-4-38

Wu, J. C., Maguire, G., Riley, G., Fallon, J., LaCasse, L., Chin, S., … Lottenberg, S. (1995). A positron emission tomography [18F]deoxyglucose study of developmental stuttering. *NeuroReport*, *6*(3), 501–505. http://doi.org/10.1097/00001756-199502000-00024

Wymbs, N. F., Ingham, R. J., Ingham, J. C., Paolini, K. E., & Grafton, S. T. (2013). Individual differences in neural regions functionally related to real and imagined stuttering. *Brain and Language*, *124*(2), 153–164. http://doi.org/10.1016/j.bandl.2012.11.013

Yairi, E. (1976). Effects of binaural and monaural noise on stuttering. *Journal of Auditory Research*, *16*(2), 114–119.

Yairi, E., & Ambrose, N. G. (1999). Early Childhood Stuttering IPersistency and Recovery Rates. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1097–1112. http://doi.org/10.1044/jslhr.4205.1097

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*(3), 213–232.

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48*(2), 607–618. http://doi.org/10.1016/j.neuropsychologia.2009.10.025

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is "where" in the human auditory cortex? *Nature Neuroscience*, *5*(9), 905–9. http://doi.org/10.1038/nn904

Zheng, Z., Munhall, K., & Johnsrude, I. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770–1781. http://doi.org/10.1162/jocn.2009.21324.Functional

Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(10), 4339–48. http://doi.org/10.1523/JNEUROSCI.6319-11.2013

Ziegler, W., Kilian, B., & Deger, K. (1997). The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia*, *35*(9), 1197–1208. http://doi.org/10.1016/S0028-3932(97)00040-7

APPENDICES

A: SYSTEMATIC REVIEW DATA EXTRACTION FORM

Study ID
Link
Intervention type
Task
Conditions
Blocks or randomised
Frequency
Duration
Control condition(s)
Participants
N
Population description
Inclusion criteria
Exclusions
Informed consent obtained
Mean age
Age range
Male
Female
Data
FMRI acquisition (TA, TR)
Corr. for multiple comparisons
Masks/ROI analysis
Statistical methods used
Results
Behavioural measure
Behavioural results
Functional results
Comparison
Co-ordinate space
Results

B: QUESTIONS USED TO ELICIT SPONTANEOUS SPEECH IN CHAPTER 4

(based on Kopelman, Wilson & Baddeley, 1989, 'The autobiographical memory interview [..]')

Can you tell me about…

Early life

- your first memory?
- A friend you had in primary school?
- A teacher you had in primary school?
- The house/area you grew up in?
- Your favourite subject when you were at school and why you liked it?
- A holiday you took as a child?

Early adult life

- A friend you had when you were a teenager?
- The first time you went on holiday on your own?
- The first time you moved house?
- Your first job?
- Someone you met at your first job?
- How your parents agreed on your names?
- One of your children's birthday parties (/a birthday party you had as a child)?

Recent life

- A relative or visitor you've seen in the last year?
- The place where you live now? (house/neighbourhood)
- Any news you've heard about a friend or relative in the last year?
- A holiday you took recently?
- Someone you met in the last year?
- A hobby you have/ what you like to do in your free time?

Culture & hobbies

- What kind of music do you enjoy? How did you get into it?
- A TV show you enjoy and what it's about
- A show you've been to at the theatre that you enjoyed
- A gig or concert you've been to that you enjoyed
- Your favourite film and what it's about
- A famous person that you admire