Error in the Superior Temporal Gyrus? A Systematic Review and Activation Likelihood Estimation Meta-Analysis of Speech Production Studies

Sophie Meekings[1] and Sophie K. Scott[2]

Abstract

■ Evidence for perceptual processing in models of speech production is often drawn from investigations in which the sound of a talker's voice is altered in real time to induce "errors." Methods of acoustic manipulation vary but are assumed to engage the same neural network and psychological processes. This paper aims to review fMRI and PET studies of altered auditory feedback and assess the strength of the evidence these studies provide for a speech error correction mechanism. Studies included were functional neuroimaging studies of speech production in neurotypical adult humans, using natural speech errors or one of three predefined speech manipulation techniques (frequency altered feedback, delayed auditory feedback, and masked auditory feedback). Seventeen studies met the inclusion criteria. In a systematic review, we evaluated whether each study (1) used an ecologically valid speech production task, (2) controlled for auditory activation caused by hearing the perturbation, (3) statistically controlled for multiple comparisons, and (4) measured behavioral compensation correlating with perturbation. None of the studies met all four criteria. We then conducted an activation likelihood estimation meta-analysis of brain coordinates from 16 studies that reported brain responses to manipulated over unmanipulated speech feedback, using the GingerALE toolbox. These foci clustered in bilateral superior temporal gyri, anterior to cortical fields typically linked to error correction. Within the limits of our analysis, we conclude that existing neuroimaging evidence is insufficient to determine whether error monitoring occurs in the posterior superior temporal gyrus regions proposed by models of speech production. ■

INTRODUCTION

A key question for investigators of speech production is to what extent, and in what fashion, we use the sound of our own voice to guide our utterances. Two widely used neural models of speech production, the hierarchical state feedback control (HSFC) model and the Directions into Velocities of Articulators (DIVA) model (Guenther & Hickok, 2015), both suggest that, in addition to prearticulatory speech monitoring, we use the sound of our own voice in error detection and correction, after speech production. Furthermore, this is achieved by a feedback circuit that compares auditory self-perception to an internal target or goal and then issues corrective signals. Both models suggest that this auditory self-perception and error correction take place in the posterior superior temporal gyrus (STG). This process of "auditory feedback control" is hypothesized to take place at the syllable level.

However, error production and correction in natural speech is unpredictable and sporadic: For example, talkers frequently do not correct mistakes (Nooteboom, 1980). To investigate the neural systems recruited to detect the errors that do occur, some functional neuroimaging studies of natural speech errors have been performed. However, these are rare, perhaps because of the difficulty eliciting such errors—to our knowledge, only two such studies have been carried out (Gauvin, De Baene, Brass, & Hartsuiker, 2016; Abel et al., 2009). Instead, most researchers wishing to investigate the mechanisms of speech error correction have relied on various methods of inducing "errors" by altering the sensory consequences of speaking aloud: The "errors" in this context reflect vocal changes associated with these alterations. This review looks at three techniques commonly used to alter the sensory consequences of speaking: frequency altered feedback (FAF), delayed auditory feedback (DAF), and masked auditory feedback (MAF). In addition, one paper in which natural speech errors were used is considered and compared to externally imposed manipulations. In this paper, and in those included in the meta-analysis, the term "feedback" refers to online auditory self-perception. Although they differ in the type of perturbation used and the assumed auditory target, these three manipulations are used to test the same hypothesis (i.e., that speech error correction occurs as described by the DIVA/HSFC models) and so must be presumed to each prompt the same error correction mechanism. From a motor control perspective, however, altering the sensory information associated with action will require modifications of the sensory control of action, which are not synonymous with speech errors. Here, we

[1]Newcastle University, [2]University College London

Journal of Cognitive Neuroscience 33:3, pp. 422–444
https://doi.org/10.1162/jocn_a_01661

conduct a systematic review of the evidence for a common error correction mechanism. Because the STG is a functionally heterogenous area (Rauschecker & Scott, 2009), activation in STG may represent different processes if the precise location of activation varies between studies. To assess the degree to which studies' results overlap and can therefore be judged to represent the same feedback control processes, an activation likelihood estimation (ALE) meta-analysis looks for convergence in reported coordinates between studies and compares these with the hypothesized location of error correction regions in the DIVA (Tourville & Guenther, 2011) and HSFC (Hickok, 2014) models.

Systematic Review

Background: Auditory Feedback Control

Any act of speech production necessarily involves movement and sound and generates somatosensory and acoustic consequences. A speech "error," therefore, might be acoustic or somatosensory. Indeed, the two are often linked; a somatosensory perturbation (e.g., talking with your mouth full) may result in acoustic inaccuracies, whereas correcting for an acoustic perturbation (e.g., raising your voice in a noisy environment) might result in somatosensory "errors." Neurally, theories of speech production (Guenther & Hickok, 2015) propose two anatomically distinct systems associated with the somatosensory control and the auditory control of speech, each with its own "feedback loop." Research suggests that talkers compensate for both somatosensory and auditory feedback perturbations and may vary in which feedback modality they prefer (Lametti, Nasir, & Ostry, 2012). Therefore, both loops are theorized to be active during speech production. However, inducing somatosensory perturbations and measuring compensation are technically complex, particularly in an MRI scanner environment where it is necessary for all equipment to be magnet-safe (Golfinopoulos et al., 2011), and as such, most neuroimaging research has focused on the auditory feedback control loop.

Figure 1 shows a simplified version of this auditory feedback control loop, as proposed by several modern models of speech production (Guenther & Hickok, 2015; Houde & Chang, 2015; Houde & Nagarajan, 2011).

During speech production, auditory control regions send a desired target to motor control regions, which issue a motor command to the vocal tract while at the same time sending a predictive, inhibitory signal to auditory control regions. Auditory feedback is processed by the sensory periphery and arrives at auditory control regions as an excitatory signal (Guenther & Hickok, 2015; Houde & Chang, 2015). If the excitatory feedback signal matches the inhibitory predictive signal (i.e., if the predicted motor act results in the predicted acoustic signal), the two will cancel each other out resulting in no net activation or even suppression. If, however, the wrong motor program was activated or auditory feedback is distorted in some way, then the signals will not be aligned and the excitatory impulse will not be inhibited, leading to a corrective signal being issued. This corrective signal results in an adjustment to the speech act—a "compensation" that corrects the error. If the same type of error persists throughout many speech acts, then the original speech motor plans may be updated, resulting in long-term "adaptation."

This is, of course, a simplified version of complex neural dynamics. The STG is a heterogenous region, and single-cell recordings in nonhuman primates have revealed distinct populations of neurons that are active or suppressed during various types of vocalization. For example, some neurons that are active when squirrel monkeys or macaques listen to vocalization playback are suppressed when vocalizing with no perturbation (Eliades & Wang, 2005, 2008; Mueller-Preuss & Ploog, 1981), but some show no difference between the

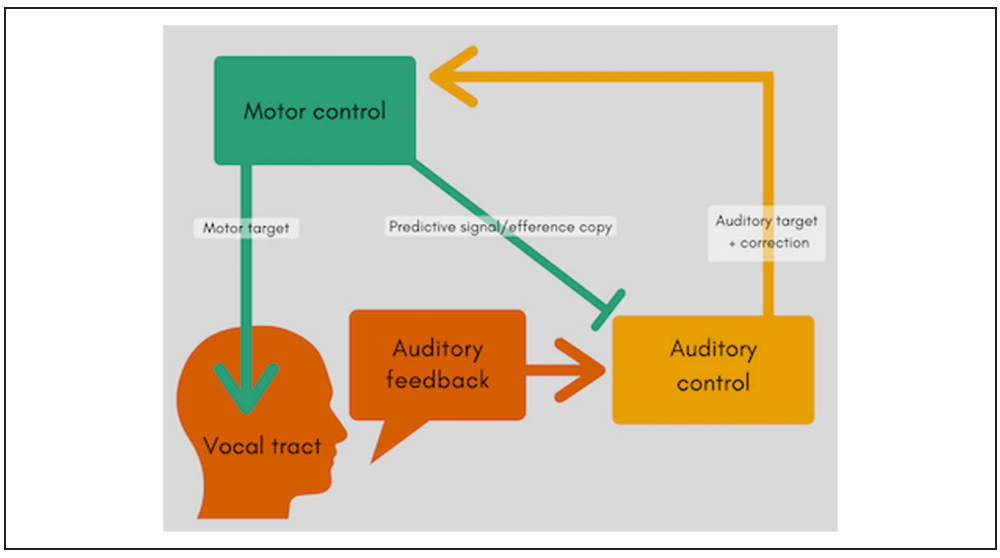

Figure 1. A simplified feedback loop of the type implemented in neural computational models of speech production. Auditory targets are defined in auditory control regions and transmitted to the motor cortex, which sends the corresponding articulatory programs to the vocal tract. Auditory feedback is compared to an efference copy or prediction from the motor cortex. If there is a match, the excitatory feedback signal is suppressed by the inhibitory predictive signal; otherwise, a corrective response is issued.

two states (Mueller-Preuss & Ploog, 1981) or are more active during vocalization than listening (Eliades & Wang, 2008). Similarly, whereas most neurons that exhibit suppression during unperturbed vocalization increase their firing rate during altered feedback, a minority decrease or do not change their firing rate (Eliades & Wang, 2008, 2012). Electrocorticography studies in humans (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Greenlee et al., 2013) have found separate populations of neurons that exhibited speaking-induced suppression, versus those that responded to altered feedback.

Activation patterns seen in humans using fMRI, which is much less spatially precise, are likely an aggregation of these different responses. In the DIVA model, three different maps in STG are proposed to reflect the activity of distinct neuronal populations—the auditory target and auditory state map, which receive excitatory input only, and the auditory error map, which can receive excitatory and inhibitory input. It is in the auditory error map (or "auditory syllable targets" region for the HSFC model) that the critical comparison of actual and predicted consequences of vocalization occurs. For the purpose of this review, we will be focusing on this "error map" as the locus of speech feedback control. The hypothesized pattern of excitation and suppression in this region should form a characteristic neural signature that is easily detected by fMRI. The four defining features of this neural signature are as follows:

1. When feedback matches what was predicted, there is (relative) suppression in auditory cortex as no corrective signal is issued (Figure 2: Normal Feedback).
2. When there is a feedback mismatch and a corrective signal is necessary, this results in increased activation in auditory control areas (Figure 2: Altered Feedback).
3. This activation should be greater in magnitude than the activation seen when listening to a comparable sound (Figure 2: Listening).
4. Increased activation in auditory control regions should be associated with corrective behavior.Figure 2 shows an example of an idealized "feedback" response, which should be accompanied by behavioral modification in altered feedback conditions according to the predictions of these models.

Overview

This review follows the PRISMA statements, an "evidence-based minimum set of items for reporting in systematic reviews and meta-analyses" (www.prisma-statement.org). The first author carried out study selection, data extraction, and evaluation of the evidence according to the following procedure. Searches using the keywords "speech auditory feedback" together with "fMRI," "magnetic resonance imaging," "PET," or "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 2017 and yielded 177 results. Forty-eight duplicates were removed, and then the remaining 129 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). Eighty-seven studies that did not meet all of these criteria were excluded at this stage. Where it was unclear whether a study met the inclusion criteria based on its abstract, the full text was read and assessed. Two studies were excluded at this stage—one that

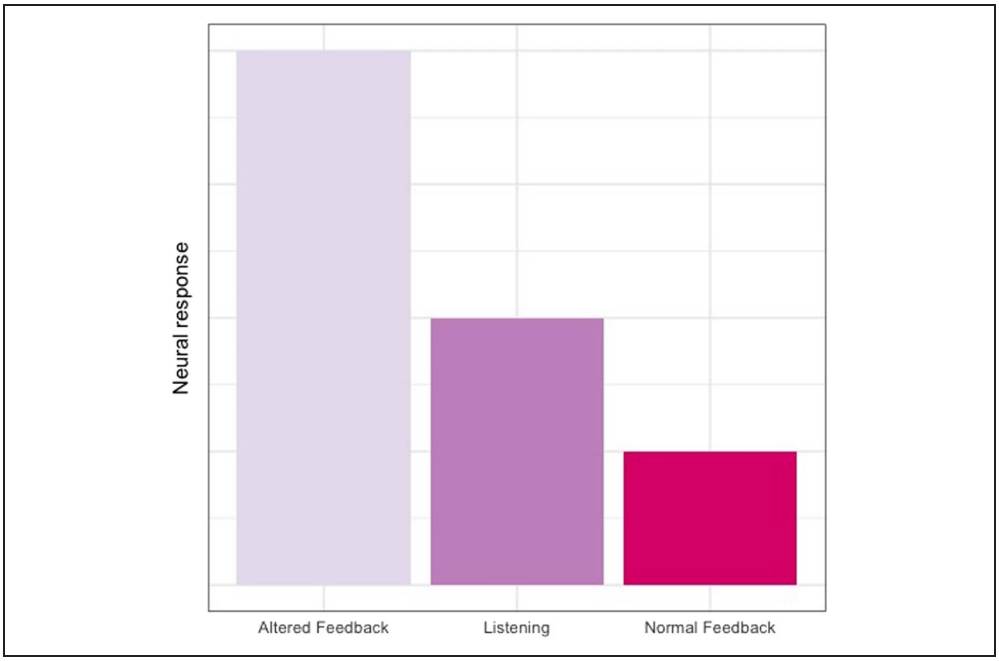

Figure 2. Example of a model "feedback- sensitive" response. Normal feedback results in reduced activation compared to listening ("speaking-induced suppression"). Altered feedback results in enhanced activation compared to listening, reflecting the corrective signal being issued.

did not include an altered feedback technique (Kell et al., 2017) and one where the experiment was designed to investigate internal, rather than external, auditory monitoring (Gauvin et al., 2016). In addition, one further study that met the criteria (Abel et al., 2009) was identified from the references of a related study. Figure 3 shows a breakdown of the study selection process. Note that, as study selection was carried out by a single person and some time has passed since the initial search, some relevant studies may have been missed or have been published since this analysis was completed.

Information was extracted from each included study as follows: (1) participants (number, inclusion criteria, age, and gender), (2) task performed (feedback alteration type, speech production task, and other experimental and control conditions), (3) neural data acquisition and analysis (acquisition parameters, stereotactic space, corrections for multiple comparisons, ROI analyses, any other statistical methods used), (4) behavioral data and results (measures of vocal compensation), and (5) neural results for the altered versus unaltered feedback comparison. A summary of the studies included is shown in Table 1.

Seventeen studies were included in the systematic review. Apart from Behroozmand et al. (2015), who used patients awaiting surgery for epilepsy, Zarate, Wood, and Zatorre (2010), who recruited singers, and Abel et al. (2009), who recruited both left- and right-handed talkers, all participants 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

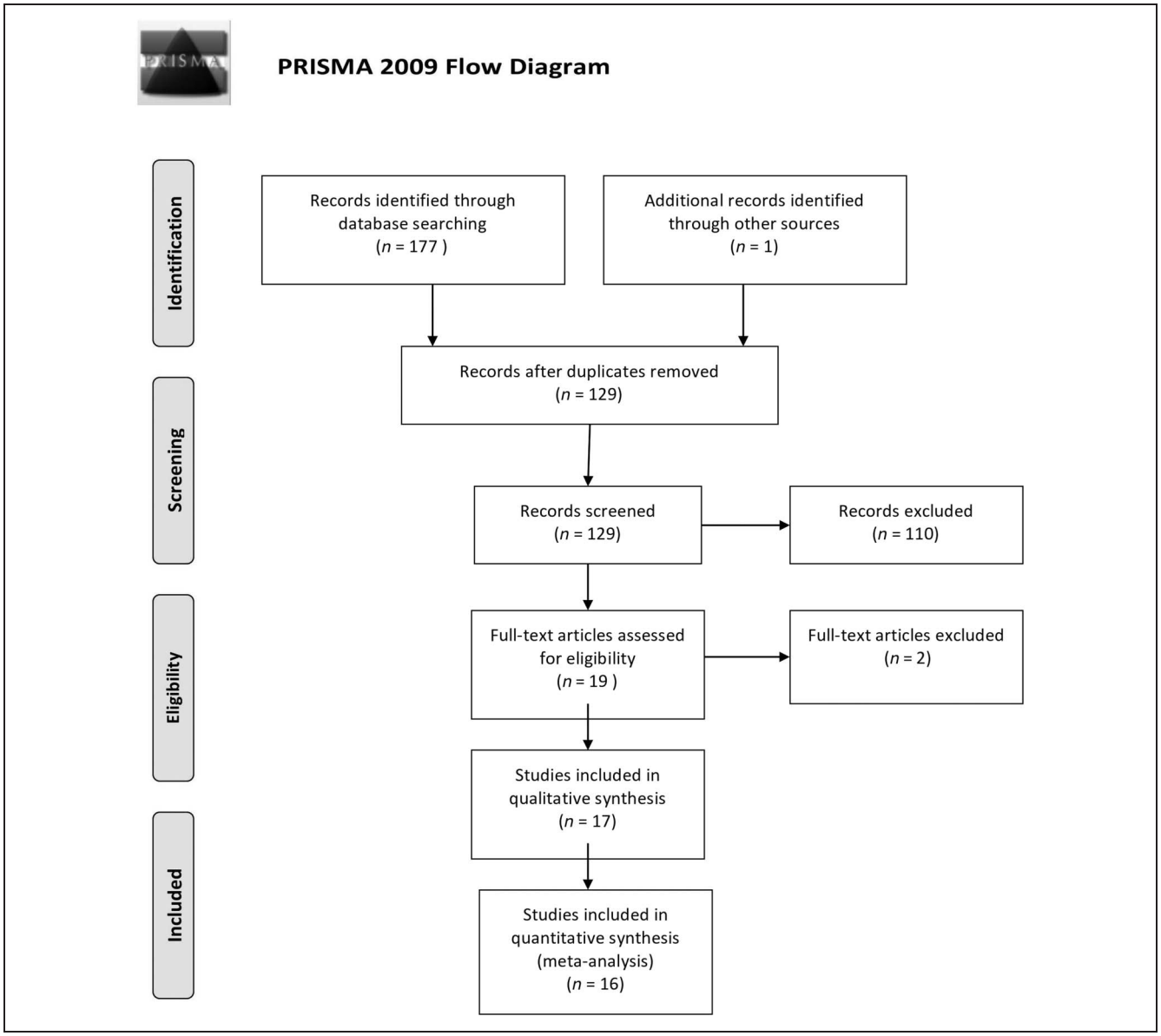

Figure 3. PRISMA flow diagram outlining study selection process.

Table 1. Summary of Studies Included in the Review, Organized Chronologically by Feedback Alteration Type

Source	*Number of Participants (After Exclusions)*	*Vocalization Task*	*Feedback Alteration*	*Population Description*	*Imaging Method*
Hirano et al. (1997)	6	Sentences	DAF	Nonclinical	PET
Hashimoto and Sakai (2003)	18	Sentences	DAF	Nonclinical	fMRI
Takaso et al. (2010)	8	Sentences	DAF	Nonclinical	PET
McGuire et al. (1996)	6	Words	FAF	Nonclinical	PET
McGuire et al. (1996)	6	Words	FAF	Nonclinical	PET
Fu et al. (2006)	13	Words	FAF	Nonclinical	fMRI
Toyomura et al. (2007)	12	Phoneme	FAF	Nonclinical	fMRI
Tourville et al. (2008)	10	Words	FAF	Nonclinical	fMRI
Zarate et al. (2010)	9	Phoneme	FAF	Singing	fMRI
Parkinson et al. (2012)	12	Phoneme	FAF	Nonclinical	fMRI
Niziolek and Guenther (2013)	15	Word	FAF	Nonclinical	fMRI
Behroozmand et al. (2015)	8	Phoneme	FAF	Epileptic	fMRI
Zheng et al. (2013)	16	Word	FAF/MAF	Nonclinical	fMRI
Christoffels et al. (2007)	14	Words	MAF	Nonclinical	fMRI
Zheng et al. (2010)	21	Word	MAF	Nonclinical	fMRI
Christoffels et al. (2011)	11	Word	MAF	Nonclinical	fMRI
Meekings et al. (2016)	14	Sentence	MAF	Nonclinical	fMRI
Abel et al. (2009)	22	Words	Natural speech errors	Nonclinical	fMRI

inclusion. In total, there were 228 participants across all 17 studies (134 men and 94 women), aged between 18 and 70 years (mean age = 31 years). Three studies used PET imaging, whereas the rest used fMRI.

Eight of 17 studies were FAF studies; three used DAF, four used noise masking (MAF), and one used both FAF and noise masking. Only one study (Abel et al., 2009) used naturally occurring speech errors instead of introducing an external perturbation. Nine studies asked participants to produce single words, either by reading aloud or by naming pictures; four studies used phoneme production, and four used whole sentences as a speech production task. The number of different stimuli (words, phonemes, or sentences) used in each experiment varied from 360 (McGuire, Silbersweig, & Frith, 1996) to only one (Zheng, Munhall, & Johnsrude, 2010).

Risk of Bias Assessment

The metafor package for R (Viechtbauer, 2015) was used to create a funnel plot as a visual indication of potential publication bias. Publication bias occurs when the result of a study influences the decision to publish it; typically, studies that find statistically significant results are more likely to be published than those with nonsignificant outcomes (Rosenthal, 1979). This is problematic as it means that the prevalence of the effect is likely to be overestimated, while contradictory evidence is effectively suppressed, meaning that published results misrepresent the underlying population effect (Scargle, 2000).

Funnel plots help assess the possibility of publication bias in the literature by plotting effect size against sample size. Studies with a smaller sample size are generally expected to show more variance in effect size, with some nonsignificant results (e.g., owing to low power) even when there is a "true" effect. This should result in a symmetrical distribution resembling an inverted funnel. However, when publication bias is present, these smaller, nonsignificant results are not reported, leading to a graph that is skewed toward the right.

T and *z* scores reported for each neural coordinate included in the meta-analysis were used to calculate Pearson's *r* as an estimate of effect size. The graph below shows the effect size for each neural point of activation plotted against the sample size of the study (Figure 4).

The plot shows a clear asymmetry, with gaps in the lower left-hand side of the plot indicating the absence of smaller

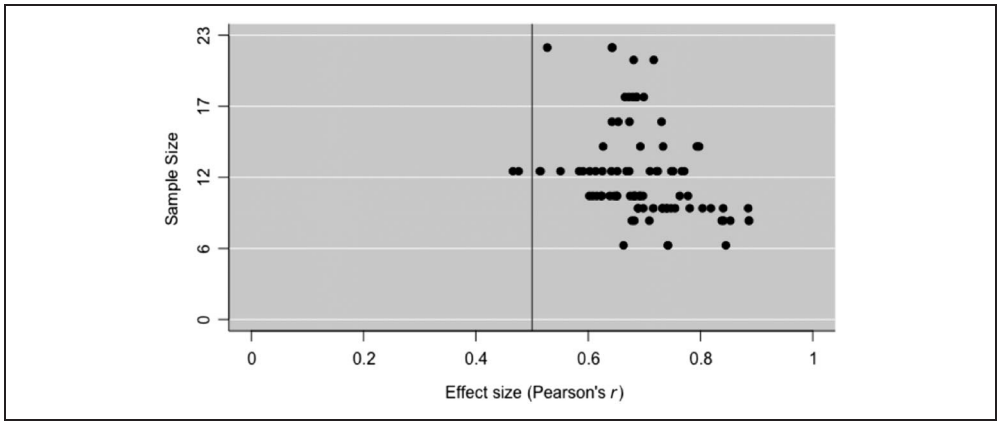

Figure 4. Funnel plot showing effect size on the x axis and sample size on the y axis, with a reference line at Pearson's $r = .5$. The expected distribution if no publication bias is present is an inverted funnel shape centered on the reference line; however, smaller studies with a low effect size are absent from the data, resulting in an asymmetrical distribution.

studies with a low effect size. Nearly all reported results had a corresponding Pearson's r of .5 or higher, indicating a moderate to very strong effect.

This suggests that publication bias may indeed be present in the data, although it should be noted that asymmetrical funnel plots may be caused by reasons other than publication bias (Sterne et al., 2011) and are not always reliably identified by visual inspection (Terrin, Schmid, & Lau, 2005). The possibility of publication bias means that the results of the meta-analysis still may not be representative of the actual pattern of activation in the underlying population.

RESULTS

The purpose of this review and meta-analysis was to establish whether the available neuroimaging evidence supports the hypothesis that the role of superior temporal cortex during speech production is that of feedback monitoring and/or error correction. Although ALE meta-analysis can estimate the overlap between results, as a mathematical technique, it is effectively blind to the methodological quality and robustness of the studies that produced those results. That is, if the input is flawed, the output will be too. To aid the interpretation of the ALE results, we conducted a systematic review of the 17 studies identified on this subject (of which 16 were taken forward to the ALE analysis), aiming to assess the overall quality of evidence. It is important to note that these were not inclusion/exclusion criteria; all studies that provided valid brain coordinates were included in the meta-analysis. Rather, the review is intended to provide with an assessment of the strengths and weaknesses of research in this area, allowing the reader to make an informed judgment about how much can be concluded from the ALE results.

We considered four central aspects of each study: task, control conditions, neural data analysis, and behavioral evidence. Studies were considered robust evidence if they used a task that reflects the way that speech is produced in everyday life, if they included a listening control condition, if they reported neural activation to altered feedback that survived correction for multiple comparisons at the peak level, and if they showed evidence of behavioral compensation that correlated with neural activation. Table 2 summarizes which studies met each criterion, and the strength of the evidence provided is discussed further below. Not meeting these criteria is not an indication of "low quality"; however, it may affect the conclusions that can be drawn from the evidence and the confidence we can have in these conclusions, both in the individual studies and at the meta-analysis level.

Did the Experiment Use a Task That Reflects the Way That Speech Is Used in Daily Life?

The nature of the speech production tasks used is important because it is possible that different levels of vocalization are processed differently and therefore recruit different brain regions. Postma (2000) describes nine different types of speech error typically seen in behavioral studies of speech production, including phonemic, prosodic, and lexical errors. It may be possible that different neural systems are associated with each type of error correction. The HSFC model (Hickok, 2014) posits that errors are detected and corrected by a motor feedback loop at the phoneme level and by auditory feedback at the syllable level, meaning that tasks that used phoneme stimuli might see more motor activation than those that used words or syllables. In addition, studies that used extended phoneme production with pitch manipulation might mimic singing, which is associated with greater activation in the right planum temporale compared to speech production (Callan et al., 2006).

In nonmanipulated speech production, there is great sensitivity to the actual task used: For example, repetition does not engage classic "Broca's area" but does engage the left anterior insula (Wise, Greene, Büchel, & Scott, 1999). As speech production tasks become more complex, differential responses are seen within left inferior frontal and posterior medial auditory fields (Blank, Scott, Murphy, Warburton, & Wise, 2002). Narrative speech has also been associated with greater activation of bilateral inferior frontal cortex compared to picture description and nonword vocalization (Troiani et al., 2008). At the sublexical level,

Table 2. Strength of Evidence Summary, Grouped Chronologically by Perturbation Type

Study	*Activation Found at Peak Level After Correction for Multiple Comparisons*	*Listening Control Condition*	*Evidence of Behavioral Compensation Linked to Neural Activation*	*Connected Speech Production Task*
Hirano et al. (1997)	No	No	No	Yes
Hashimoto and Sakai (2003)	Unclear	No	Yes	Yes
Takaso et al. (2010)	No	No	Yes	Yes
McGuire et al. (1996)	No	Yes	No	No
Fu et al. (2006)	No	No	No	No
Toyomura et al. (2007)	Unclear	No	No	No
Tourville et al. (2008)	No	No	Yes	No
Zarate et al. (2010)	Yes	Yes	Yes	No
Parkinson et al. (2012)	No	No	No	No
Niziolek and Guenther (2013)	Yes	No	Yes	No
Behroozmand et al. (2015)	No	Yes	Yes	No
Zheng et al. (2013)	Yes	Yes	No	No
Christoffels et al. (2007)	No	Yes	No	No
Zheng et al. (2010)	Unclear	Yes	No	No
Christoffels et al. (2011)	No	Yes	No	No
Meekings et al. (2016)	Yes	Yes	No	Yes
Abel et al. (2009)	No	No	Unclear	No

phoneme production, syllable production, and phonological chunking processes recruit spatially distinct networks (Peeva et al., 2010).

Because models such as DIVA are intended to reflect the mechanisms by which speech is produced in daily life, ideally experiments testing these models would use naturalistic speech production tasks—for example, connected speech rather than single phoneme production. However, the choice of task is necessarily the result of an interplay between sparse fMRI scanning constraints (which require the utterance to be relatively short to fit in the silent gap between scans) and feedback alteration type. Consequently, 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.

Compensation to FAF is strongest when perturbation duration exceeds 100 msec (Burnett, Freedland, Larson, & Hain, 1998) and so requires extended vocalization to prompt behavioral compensation, as a typical articulation rate in conversational speech is 10 phonemes per second (Osser & Peng, 1964)—that is, on average, one phoneme will take around 100 msec to produce. Thus, four FAF studies required participants to articulate the phoneme /a/ for up to 5 sec, dozens of times per experiment. Some other studies that used single words also required talkers to prolong their utterances. Tourville et al. (2008) trained participants to produce each word more slowly than usual, resulting in a mean vowel duration of between 357 and 593 msec. Niziolek and Guenther (2013) found compensation for formant shifting 400 msec from speech onset, suggesting that participants spoke slowly enough to allow the formant manipulation to be applied. These studies are thus using speech production over far greater durations than seen in normal speech: Producing a single phoneme over 5 sec is 50 times slower than typical articulation. Indeed, this is closer to singing than to speech, in that it requires extended breath control—and in fact, Zarate et al. (2010) used such production as a singing task.

In studies that used single word production, the stimuli set was often highly restricted, with as few as eight different monosyllabic word stimuli, and one study (Zheng et al., 2013) used only 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, the psychological phenomenon in which continuous repetition of a word causes it to lose meaning for the speaker and instead be perceived as a gibberish sound (Smith & Klein, 1990). Although it appears that semantic content is not necessary to prompt adaption, because 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); the suppression of neural responses is a well-established

Table 3. A Breakdown of the Tasks Used by Each Study, Grouped Chronologically by Perturbation Type

Study	*"Error" Type*	*Perturbation Magnitude*	*Specific Stimulus Type*	*Number of Discrete Stimuli*	*Number of Vocalization Trials*	*Stimulus Repetitions*
Hirano et al. (1997)	DAF/low-pass filter	100 msec/filter cutoff = 300 Hz	"Familiar" sentences	17	34	2
Hashimoto and Sakai (2003)	DAF	200 msec	Seven-syllable sentences	27	864	32
Takaso et al. (2010)	DAF	50, 125, and 200 msec	Passages from a children's book	16	16	1
McGuire et al. (1996)	FAF	800 cents up	Monosyllabic or bisyllabic nouns	360	360	1
Fu et al. (2006)	FAF		Adjectives	96	96	1
Toyomura et al. (2007)	FAF	200 cents, up or down	Single vowel /a/	1	120	120
Tourville et al. (2008)	FAF (F1 manipulation)	Shifted up or down by 30%	Monosyllabic CVC words	8	256	32
Zarate et al. (2010)	FAF	200 and 25 cents, up or down	Single vowel /a/	1	80	80
Parkinson et al. (2012)	FAF	100 cents, up or down	Single vowel /a/	1	144	144
Niziolek and Guenther (2013)	FAF (F1 and F2 manipulation)	Subjectively determined for each participant	Monosyllabic CVC words	8	320	40
Behroozmand et al. (2015)	FAF	600 cents up	Single vowel /a/	1	80	80
Zheng et al. (2013)	FAF (F1 and F2 manipulation)/ MAF (signal-correlated noise)	F1 and F2 shifted by 200 and 250 Hz, respectively; noise played at 85 dB SPL	Monosyllabic words	2	216	108
Christoffels et al. (2007)	MAF (pink noise at three different intensities)	Max level subjectively determined for each participant and then decreased by 10 and 15 dB SPL to create two further maskers	Monosyllabic or bisyllabic nouns	25	150	6
Zheng et al. (2010)	MAF (signal-correlated noise)	85 dB SPL	Monosyllabic CVC words	1	216	216
Christoffels et al. (2011)	MAF (pink noise)	Subjectively determined for each participant	Monosyllabic or bisyllabic nouns	20	60	3
Meekings et al. (2016)	MAF (white noise, signal-correlated noise, rotated speech, speech)	84 dB SPL	Short sentences, maximum = 7 syllables	200	200	1
Abel et al. (2009)	Natural speech errors	–	Monosyllabic nouns	132	132	1

"'Error' type" and "perturbation magnitude" columns describe what each participant heard, whereas the remaining columns describe the type of speech production elicited. CVC = consonant - verb - consonant.

consequence of repeated stimulus presentation in several domains (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Henson & Rugg, 2003; Desimone, 1996). Only four studies used connected speech (sentences) as stimuli (Meekings et al., 2016; Takaso, Eisner, Wise, & Scott, 2010; Hashimoto & Sakai, 2003; Hirano et al., 1997). Two of the three DAF studies used PET, in which continuous scanning during speech is possible with no scanner noise and fewer movement artifact issues so there are fewer constraints on task length. However, sparse fMRI scanning allows a silent period of approximately 3 sec in which it is possible to speak while minimizing movement artifacts (Gracco, Tremblay, & Pike, 2005), so although it is not possible for participants to read entire passages, sentence production is certainly possible.

The nature of the manipulation used is also of interest when assessing experiments' ecological validity. Because the aim of this 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. Although Abel et al.'s (2009) picture-naming task failed to elicit many phonetic errors, paradigms that successfully evoke natural speech errors without external manipulation are likely to come closest to "speech error" as conceptualized by speech feedback models. However, most studies used more artificial or noticeable manipulations such as DAF or pitch shift that affected the whole utterance. Also notable is that, even among studies that used the same type of manipulation, there was very little coherence in protocol between studies, with each investigation varying considerably in perturbation magnitude and direction (where relevant). That is, there appears to be no strong consensus on what constitutes an auditory error nor how best to evoke it.

Masking noise might be considered the most ecologically valid approach, because most people will experience a conversation in a noisy environment outside the laboratory, whereas they are unlikely ever to hear the pitch of their voice shift suddenly unless they make a habit of inhaling helium, and DAF is rarely heard outside faulty phone lines or recording booths. Similarly, spectral modulations that affect only the first and second formants (as used by Niziolek & Guenther, 2013; Zheng et al., 2013; Tourville, Reilly, & Guenther, 2008) mimic the kind of misarticulating that the error correction mechanism is supposed to deal with, with both the HSFC (Hickok, 2014) and DIVA (Tourville & Guenther, 2011) models stating that auditory targets are likely defined at the syllable level.

Controls: Did the Study Include a Listening Control Condition?

When we speak normally, without perturbation or error, auditory cortex typically displays suppression of activation compared to when we hear sounds that are not self-generated (Flinker et al., 2010; also Houde et al., 2002; and Wise et al., 1999). This is called the "speaking-induced suppression" effect, and it is part of the three-way activation profile shown in Figure 2. Demonstrating speaking-induced suppression is useful as a way of showing that auditory feedback control is going on—the DIVA and HSFC models hypothesize that suppression occurs when no error is detected, and auditory and motor control signals cancel each other out. That is, speaking-induced suppression is theorized to occur as the result of the "same" mechanisms that cause error detection and correction, and the same brain regions that are suppressed in activity during normal, error-free speech production are also those that should show an enhanced response when speech errors are produced.

In addition to allowing us to identify speaking-induced suppression and thus build the profile of the feedback-sensitive region, including a condition in which participants listen to sounds without vocalizing or articulating is also important as a control for the auditory component of feedback manipulation. The three perturbation techniques used in most studies all involve playing additional sounds over headphones. Although the additional sound, in the case of DAF and FAF (but not MAF), is the participant's own voice, perception of our own self-generated speech (which we hear through air and bone conduction) is different to perception of air-conducted manipulated speech played through headphones (Pörschmann, 2000).

When speech is unmanipulated, talkers' abilities to accurately identify recordings of their own voice heard through air conduction vary considerably, with accuracy rates from as low as 38% (Rousey & Holzman, 1967) to as high as 96% (Rosa, Lassonde, Pinard, Keenan, & Belin, 2008). Introducing a perturbation, such as pitch-shifting an utterance, significantly decreases the likelihood of talkers recognizing the utterance as self-generated (Allen et al., 2005). In addition, when participants hear manipulated speech through headphones, they are frequently able to consciously identify that they are being played a manipulation (Meekings et al., 2015; Hafke, 2008; Elman, 1981) rather than hearing themselves. This means that we cannot simply take for granted that participants are processing feedback perturbations as self-generated speech and not an irrelevant masker. A listening control condition allows us to factor out activation associated simply with perceiving new sounds and interrogate whether there is an additional effect of error detection. Without this control, any enhanced activation in altered feedback compared to speaking could be attributed to the fact that participants are hearing a novel sound during altered auditory feedback (rather than specifically interpreting it as an error in their speech).

However, more than half of the studies discussed here did not include a listening condition, so we were unable to confirm the presence of speaking-induced suppression or rule out that enhanced STG responses to altered feedback were simply a response to hearing something unusual. Notably, the only study that did not introduce an external perturbation (Abel et al., 2009) also found no activation in STG.

Table 4. Details of the Neural Analysis Carried Out by Each Study (Grouped Chronologically by Perturbation Type) and the Regions in which Activation Was Found in Response to the Specified Contrast

Study	*Feedback Alteration*	*Contrast Used*	*Threshold*	*Number of Foci at the Whole-Brain Level*	*Additional Analyses*	*Regions in which Activation Found*
Hirano et al. (1997)	DAF	Delay > rest	Uncorrected	10		Bilateral Heschl's gyrus, STG, Broca's area, motor area, cerebellum, visual cortices
Hashimoto and Sakai (2003)	DAF	Delay > normal feedback	$p < .05$, corrected (correction not specified)	6		Bilateral STG, SMG, MTG
Takaso et al. (2010)	DAF	Parametric response to delay increase	Uncorrected $p < .0001$	5		Bilateral STG
McGuire et al. (1996)	FAF	Pitch shift > normal feedback	Uncorrected $p < .001$	4		Bilateral STG, R STS, L insula
Fu et al. (2006)	FAF	Pitch shift > normal feedback	Cluster level $p < .01$	21		Bilateral STG, ACC, posterior cingulate, right IFG, primary occipital cortex, putamen, and brainstem
Toyomura et al. (2007)	FAF	Pitch shift > normal feedback	Corrected $p < .05$	5		R supramarginal gyrus, pFC, anterior insula, STG/STS, L premotor area
Tourville et al. (2008)	FAF	Pitch shift > normal feedback	FDR $p < .05$	None	Fixed effect analysis and 142 ROIs based on speech model predictions	Bilateral posterior STG and planum temporale; R vMC, vPMC, and amCB
Zarate et al. (2010)	FAF	Conjunction of two pitch shift conditions > normal feedback	FWE $p < .05$	15	Small-volume correction for regions that fell below threshold but had been significant in a previous study	Bilateral BA 6/55, anterior insulae, pre-SMA, right RCZa, bilateral mid-PMC, intraparietal sulci, supramarginal gyri, right STS, and planum temporale
Parkinson et al. (2012)	FAF	Pitch shift > normal feedback	Uncorrected $p < .001$	6		Bilateral STG
Niziolek and Guenther (2013)	FAF	Pitch shift > normal feedback	FDR $p < .05$	8		Bilateral posterior STG, bilateral IFG, pars opercularis, and pars triangularis

Table 4. (*continued*)

Study	*Feedback Alteration*	*Contrast Used*	*Threshold*	*Number of Foci at the Whole-Brain Level*	*Additional Analyses*	*Regions in which Activation Found*
Behroozmand et al. (2015)	FAF	Pitch shift > normal feedback	FWE $p < .05$	None	8 ROIs based on shift/no shift vs. rest functional maps	Bilateral STG and precentral gyri
Zheng et al. (2013)	FAF/MAF	MVPA (pitch shift and MAF) > (pitch shift and no shift) AND (MAF and no shift)	FWE $p < .05$	4		Bilateral cerebellum, right angular gyrus, and right SMA
Christoffels et al. (2007)	MAF	Noise > normal feedback	Cluster level $p < .05$	3		Bilateral STG
Zheng et al. (2010)	MAF	Interaction between MAFed > normal feedback and listening > normal feedback	$p < .05$, corrected (correction not specified)	2		Bilateral STG, inferior STS and MTG
Christoffels et al. (2011)	MAF	Parametric response to noise MAFing	FWE $p < .05$	None	ROI mask based on uncorrected contrast of speaking in quiet < speaking in the loudest noise level	Bilateral STG
Meekings et al. (2016)	MAF	Listening > MAF > no MAF	FWE $p < .05$	2		Bilateral STG
Abel et al. (2009)	Natural speech errors	Incorrect > correct picture naming	Uncorrected peak $p < .001$, cluster level FWE $p < .05$	None/3		R SMA and MFG, L insula

BA = Brodmann's area; FDR = false discovery rate; IFG = inferior frontal gyrus; L = left; MTG = middle temporal gyrus; R = right; amCB = anterior medial cerebellum; MVPA = multi-voxel pattern analysis; RCZa = anterior rostral cingulate zone; PMC = premotor cortex; SMG = supramarginal gyrus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.

Table 5. Behavioral Data Collected by Each Study (Grouped Chronologically by Perturbation Type)

Study	*Behavioral Compensation Reported*	*Other Behavioral Analysis/Results*	*Adaption Magnitude as Percentage of Perturbation Magnitude*	*Compensation Accompanied by Neural Response in STG*
Hirano et al. (1997)	Fluency decrease in DAF anecdotally reported but no quantitative evidence presented	N/A	–	–
Hashimoto and Sakai (2003)	Fluency (measured in correctly spoken morae per second)	Participants were less fluent in the DAF condition	–	–
Takaso et al. (2010)	Likert scale ratings for perceived speed, difficulty, and accuracy of articulation	As delay increased, so did perceived difficulty, whereas speed and accuracy ratings decreased.	–	–
McGuire et al. (1996)	Pitch change anecdotally reported but no quantitative evidence presented	Participants asked to attribute their speech to themselves or another; made no misattributions	–	–
Fu et al. (2006)	Participants asked to attribute their speech to themselves or another	Participants made more attribution errors in FAF compared to unaltered feedback	–	–
Toyomura et al. (2007)	Pitch change anecdotally reported but no quantitative evidence presented	–	–	–
Tourville et al. (2008) (shift down)	Pitch change	–	13	Yes
Tourville et al. (2008; shift up)	Pitch change	–	13.6	
Zarate et al. (2010; 200-cent shift, deliberate compensation)	Pitch change	–	88	Yes
Zarate et al. (2010; 200-cent shift, suppressing response)	Pitch change	–	10	
Zarate et al. (2010; 25-cent shift, deliberate compensation)	Pitch change	–	113	
Zarate et al., (2010; 25-cent shift, suppressing response)	Pitch change	–	72	

Table 5. (*continued*)

Study	*Behavioral Compensation Reported*	*Other Behavioral Analysis/Results*	*Adaption Magnitude as Percentage of Perturbation Magnitude*	*Compensation Accompanied by Neural Response in STG*
Parkinson et al. (2012; down shift)	Pitch change	–	21.36	
Parkinson et al. (2012; up shift)	Pitch change	–	17.47	Yes
Niziolek and Guenther (2013; far from phoneme boundary)	Pitch change	–	3	Yes
Niziolek and Guenther (2013; near phoneme boundary)	Pitch change	–	25	
Behroozmand et al. (2015)	Pitch change	–	4.5	Yes
Zheng et al. (2013)	None reported	–	–	–
Christoffels et al. (2007)	Participants asked to suppress compensation response	–	–	–
Zheng et al. (2010)	None reported	–	–	–
Christoffels et al. (2011)	Participants asked to suppress compensation response	–	–	–
Meekings et al. (2016)	Vocal intensity	–	56	No
Abel et al. (2009)	Picture naming errors	7% error rate; mainly semantic	–	No

Where quantitative compensation data was available, this has been used to calculate the mean compensation response as a percentage of the perturbation manipulation, allowing direct comparison of the degree of compensation found in each study.

Of those that did include a listening condition, three (Behroozmand et al., 2015; Christoffels, Formisano, & Schiller, 2007; McGuire et al., 1996) 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 four studies (Meekings et al., 2016; Zheng et al., 2010, 2013; Christoffels, van de Ven, Waldorp, Formisano, & Schiller, 2011) found evidence of speaking-induced suppression, whereas 12 either found no evidence or did not look for it.

Analysis: Was Significant Activation Found at Peak, Whole-Brain Level After Correction for Multiple Comparisons?

Leaving the interpretability of the results to one side for the moment, it is also important to evaluate the robustness of the statistical analysis that produced them. Because brain data analysis involves testing for activation at hundreds of thousands of voxels across the brain, it is very likely that some voxels will appear active just by chance. For this reason, it is necessary to apply a statistical correction for multiple comparisons. Historically, brain imaging studies have attempted to control the false-positive rate by setting a high, if uncorrected, significance threshold, such as $p <$ .001. However, research has demonstrated that this is insufficient to prevent spurious activation from appearing (Bennett, Baird, Miller, & Wolford, 2004).

Data can be corrected for multiple comparisons at the cluster level (i.e., a group of voxels is significantly more active in one condition than another) or at the peak level (a single voxel is significantly more active in one condition than another). Owing to an error of implementation in several commonly used software packages, cluster-level correction has been reported to result in false-positive rates of up to 70% (Eklund, Nichols, & Knutsson, 2016). We therefore considered that the best evidence for significant activation came from studies that used correction for multiple comparisons at the peak level, although we did not exclude any study from the ALE meta-analysis on the basis of the statistical analysis applied.

Of the 17 studies considered, five used an uncorrected peak threshold (Parkinson et al., 2012; Takaso et al., 2010; Abel et al., 2009; Hirano et al., 1997; McGuire et al., 1996). Three used cluster-level correction (Abel et al., 2009; Christoffels et al., 2007; Fu et al., 2006). Three studies (Zheng et al., 2010; Toyomura et al., 2007; Hashimoto & Sakai, 2003) used a threshold of $p < .05$ corrected for multiple comparisons but did not provide specific information on whether this correction was applied at the peak or cluster level or on the method used (such as FWE or false discovery rate). Seven studies used peak-level correction for multiple comparisons, but of those, three found no significant activation when this correction was applied (Behroozmand et al., 2015; Christoffels et al., 2011; Tourville et al., 2008), although all three of these studies went on to report a positive result using less stringent methods, such as fixed effects analysis (Tourville et al., 2008) or uncorrected p thresholds (Behroozmand et al., 2015), or by restricting the search to auditory cortex and similar ROIs (Christoffels et al., 2011; Tourville et al., 2008). Note that these methods also affect the interpretability of the results: Fixed effects analyses cannot be used to generalize beyond the group tested, whereas determining ROIs after a fixed effects analysis has confirmed where activation can be found may artificially increase the likelihood of finding significant results (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

In summary, only 4 of the 17 studies can be positively identified to have found a neural response to manipulated, contrasted with unmanipulated, speech production at the most stringent significance threshold (Meekings et al., 2016; Niziolek & Guenther, 2013; Zheng et al., 2013; Zarate et al., 2010).

Uncorrected p thresholds are relatively common in older (pre-2010) neuroimaging studies, and issues with false positives in some methods of cluster-level corrections have only recently been reported on. In addition, it is possible that the "feedback correction" effect is relatively small, and studies did not have enough power to detect an effect because of small sample size. We therefore did not exclude any study from the meta-analysis on the basis of the method of correction used, and we included coordinates from studies that initially found null results at the peak level.

Behavior: Did the Authors Report Behavioral Evidence that Correlates with Neural Activation?

Acoustic data can help to interpret the neural response by showing if compensation has taken place. If a neural response is associated with feedback control, then it should be followed by a vocal correction for the "error." Table 5 shows these data for studies that reported it. Seven of 17 studies, consisting of five FAF studies, one MAF study, and one DAF study, reported the results of acoustic analyses. FAF studies typically reported response direction (in the same or opposite direction to the shift), magnitude (how much participants compensated), and latency (time taken to compensate for the shift), 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 200 msec of stimulus onset, which is longer than the duration of a phoneme at typical speech rates but approximately the average duration of a syllable (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: That is, people do not adapt their speech to compensate very much for the shift in perceived pitch, 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. Other studies have found that human participants tend not to reliably or fully compensate for changes in pitch (Chang et al., 2013; Katseff, Houde, & Johnson, 2012) or loudness (Lane & Tranel, 1971). However, when asked to deliberately compensate for a frequency shift, Zarate et al. (2010) found that participants were capable of compensating for 87.6% of a 200-cent frequency shift and overcompensated for a 25-cent shift, suggesting that talkers are capable of fully or nearly fully compensating for pitch shift despite the perceptual difficulties outlined above. Although their participants were trained singers and were therefore likely to perform better at this task than nonsingers, nonsingers are also able to make considerable adjustments when prompted to do so (Murry, 1990).

Figure 5 shows the magnitude of behavioral responses as a percentage of the total perturbation magnitude.

In summary, then, evidence from the seven studies that reported acoustic analysis suggests that behavioral responses are relatively small and do not fully compensate for perturbations. This is the case even when talkers are trained to speak slowly or sustain their vocalizations to maximize response to the perturbation (Behroozmand et al., 2015; Parkinson et al., 2012; Zarate et al., 2010; Tourville et al., 2008).

Two studies (Christoffels et al., 2007, 2011) explicitly instructed participants to avoid any kind of compensation 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, the cognitive effort involved in this suppression may confound the neural results as this involves suppressing the natural behavioral response to noise. A similar masking noise study (Meekings et al., 2016), which allowed talkers to modulate their voices naturally, found no brain regions in which activation positively correlated with behavioral compensation.

Activation Likelihood Meta-Analysis: Methods

To assess the convergence in "feedback" regions across studies, an ALE analysis was carried out. Peak coordinates resulting from the altered > unaltered speech manipulation contrasts in each study were selected for inclusion, unless a study used an analysis that explicitly compared altered feedback with both unaltered feedback and listening (e.g., Zheng et al.'s [2010] interaction analysis). In that case, those coordinates were used instead; the aim was to select coordinates 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.

Coordinates given in Talairach space were converted to Montreal Neurological Institute (MNI) space for the meta-analysis, using the Brett transform (Brett, Christoff, Cusack, & Lancaster, 2001). A visualization of each of the points included in the meta-analysis was created by building 1-mm

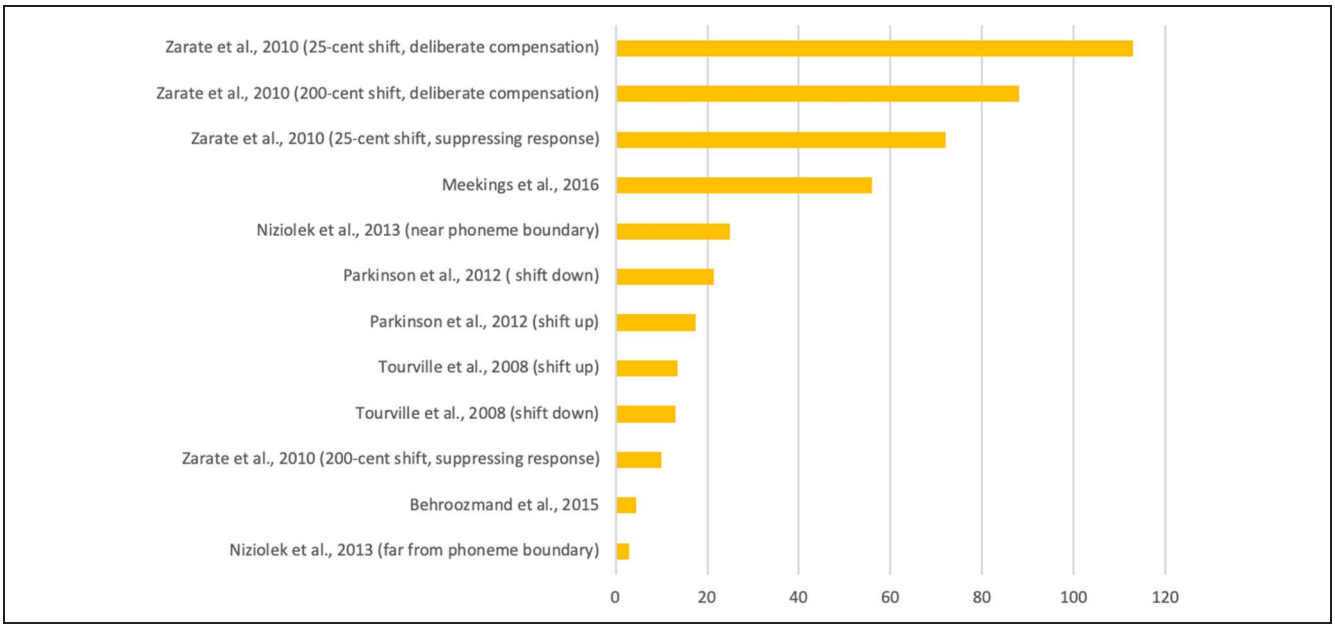

Figure 5. Compensation as a percentage of the perturbation magnitude, for the studies that reported this data. Shown ranked from most compensation to least compensation. All studies were FAF studies apart from Meekings et al. (2016), which used MAF.

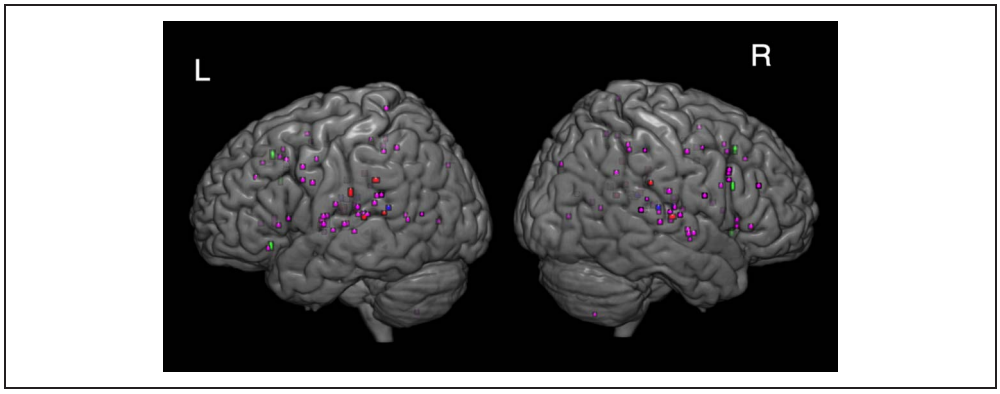

Figure 6. All coordinates entered into the ALE analysis, converted to MNI coordinates and plotted as spheres with a 1-mm radius. View from the surface of the brain. Purple = FAF, red = DAF, blue = MAF, and green = natural speech errors. Fainter dots represent activation below the surface of the brain.

spherical ROIs for each coordinate using MarsBaR software (Brett, Anton, Valabregue, & Poline, 2002) using a custom script to generate color-coded ROI maps showing all the coordinates for each perturbation type. These maps were projected onto the Colin 27 average brain (Holmes et al., 1998), and the result is shown in Figure 6. This allows a visual comparison between different perturbation types, as the sample size was too small to create separate ALE maps for each study type.

An ALE meta-analysis was carried out using a nonadditive random effects model, as described by Eickhoff et al. (Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012; 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 modeling each coordinate using a 3-D Gaussian probability density function with an FWHM determined by the number of participants in each study (median FWHM = 9.7 mm, range = 9.2–10.9 mm). 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 $p < .05$, recommended by Eickhoff et al. (2016) as the least likely to result in false positives with a sample size of fewer than 17 studies. There was no minimum cluster size. The results were projected onto a standard template in MNI space (Holmes et al., 1998).

ALE Meta-Analysis: Results

The ALE analysis revealed four significant clusters, two in the right hemisphere and two in the left hemisphere (Figures 7 and 8). In both hemispheres, activation spanned superior and transverse temporal gyri, whereas in the right hemisphere, activation also encompassed precentral and postcentral gyri as well as insula. Eighteen foci from 12 studies contributed to the first cluster, in the right hemisphere. Fewer studies overlapped in the left hemisphere, with six foci from five different studies contributing to two significant clusters. Cluster extent and coordinates of ALE peaks are given in Table 6, along with probabilistic anatomical labels for each extremum derived automatically by the GingerALE software. In the right hemisphere, there were four peaks: two in posterior STG, one in the transverse temporal gyrus, and one in the precentral gyrus. In the left hemisphere, two peaks were found in posterior STG.

To compare results to projected auditory error/target regions, 3-mm spherical ROIs were created around the hypothesized neural correlates of the DIVA auditory target/error maps as defined by Golfinopoulos, Tourville, and Guenther (2010) using the MarsBaR toolbox (Brett et al., 2002). These ROIs were superimposed onto the ALE map showing regions of significant convergence between studies. The two maps did not overlap, although activation in the right posterior STG bordered on the DIVA region of interest based around coordinates [69.5, 30.7, 5.2], as shown in Figure 9.

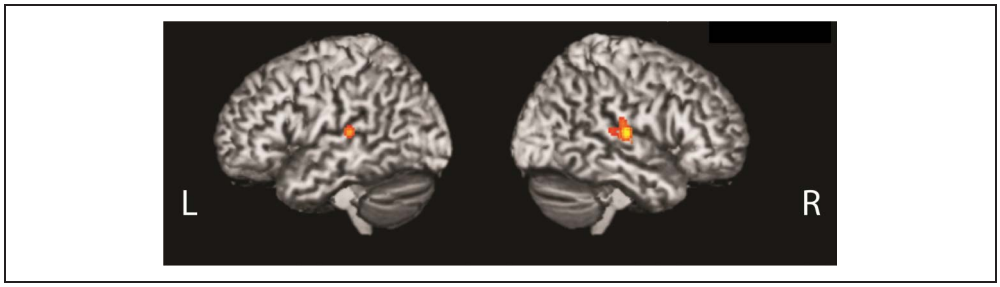

Figure 7. ALE map, view from the surface of the brain. This image shows regions of significant convergence between activation foci in the 14 selected auditory feedback perturbation studies, as revealed by an ALE analysis. Corrected for multiple comparisons using peak-level FWE $p < .05$.

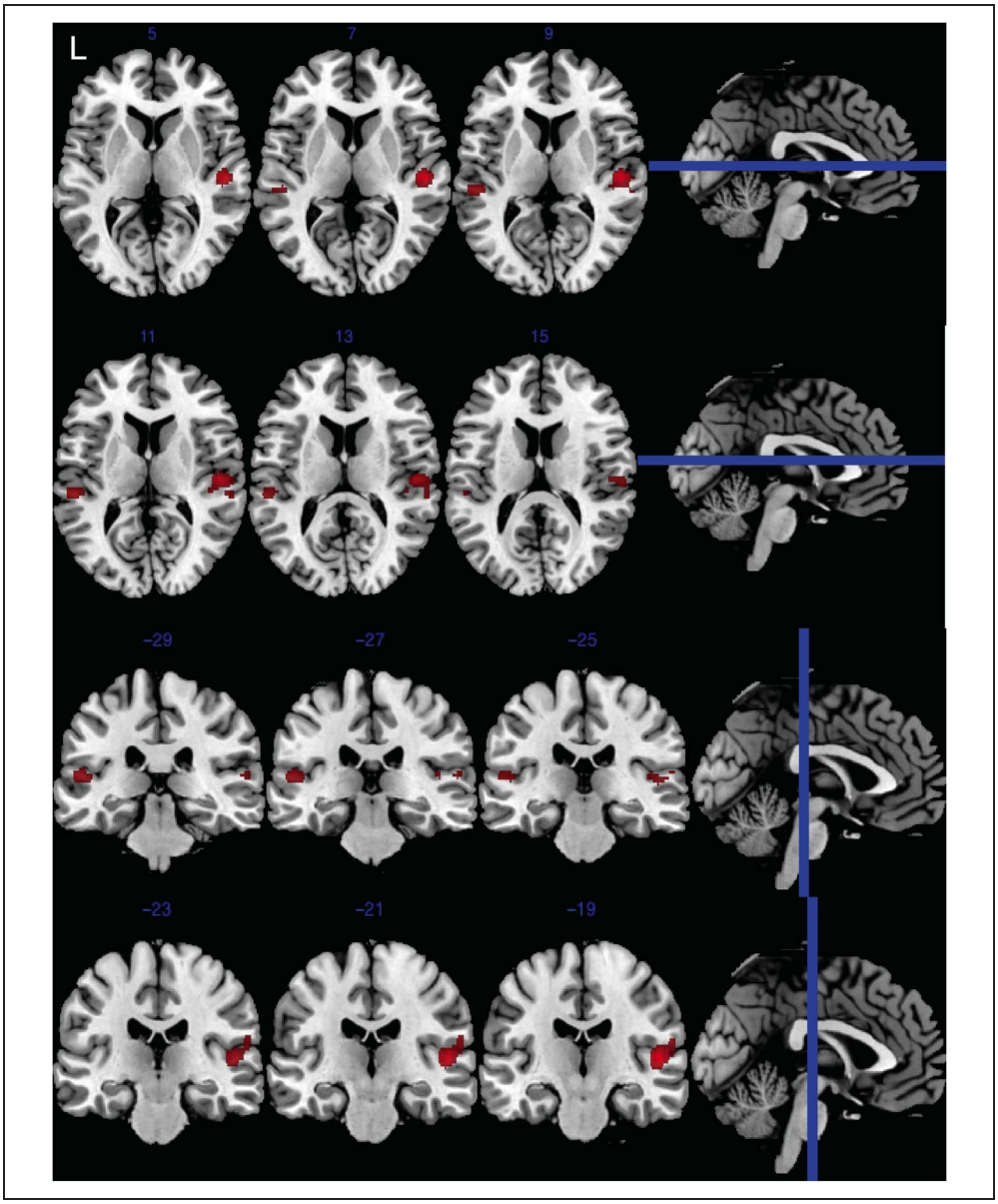

Figure 8. ALE map, multislice view. Regions of significant convergence between studies, corrected for multiple comparisons at peak-level FWE $p < .05$.

In addition, we investigated the possibility that activation seen was related to auditory–motor transformation as suggested by the HSFC (Hickok, 2014). Area Spt, a functionally defined area located in the left hemisphere at the parietal–temporal boundary within the lateral sulcus (Buchsbaum, Olsen, Koch, & Berman, 2005), is hypothesized to convey information from auditory target regions to motor areas where a corrective signal is issued. To gain an estimate of potential overlap between the ALE results and Spt, a 3-mm sphere ROI was created around the mean location of Spt in multiple single-participant analyses as reported by Hickok, Okada, and Serences (2009), converted from Talairach into MNI space. Figure 10 shows both the Spt ROI (pink) and the ALE results (red/yellow) superimposed on a standard brain in MNI space.

Again, there was no overlap between the estimated location of Spt and areas of significant convergence between studies as revealed by the ALE. ALE activation was more anterior than both Spt and DIVA auditory target/error regions.

Summary: ALE Analysis

The ALE analysis found significant overlap between experimental foci in STG, transverse temporal gyrus, and precentral gyrus. In STG, overlap was concentrated in posterior regions and was more widespread in the right than the left hemisphere. This activation was anterior to ROIs identified as critical to auditory error correction (Golfinopoulos et al., 2010), although in one case, the DIVA and ALE maps bordered on each other, suggesting that some overlap may occur if a different comparison method was employed. Meanwhile, an ROI proposed as the site of auditory–motor transform (Hickok et al., 2009)

Table 6. Regions of Significant Convergence between Studies, as Revealed by the ALE Meta-Analysis

Cluster #	*Foci (Studies)*	*Volume (mm^3)*	*Weighted Center (x, y, z)*			*Peak Value*	*Peak Coordinates (x, y, z)*			*Hemisphere*	*Macroanatomical Label*	*Cytoarchitectonic Label*
1	16 (10)	2296	52.9	−20.8	8.4	0.0376	52	−20	8	Right	Transverse temporal gyrus	BA 41
						0.0210	60	−22	16	Right	Postcentral gyrus	BA 40
						0.0189	58	−30	10	Right	STG	BA 41
2	6 (5)	648	−52.6	−28.8	9.5	0.0239	−54	−30	10	Left	STG	BA 41
3	1 (1)	8	52	−10	−2	0.0164	52	−10	2	Right	STG	BA 22
4	0	8	−50	−16	2	0.0160	−50	−16	2	Left	STG	BA 22

was considerably posterior to and spatially distinct from the ALE activation sites.

As summarized above, the quality of the evidence provided by the studies included in the meta-analysis was somewhat variable. Neural responses varied in strength across studies, with many experiments failing to find results at the whole-brain level when corrected for multiple comparisons. By pooling results, meta-analysis can ameliorate issues of low statistical significance; an effect that does not reach significance in an isolated study may well reflect a real effect that becomes significant when results are aggregated. A more serious problem is that many studies did not include a listening control condition, to rule out the possibility that activation was related to hearing unusual sounds in the manipulated feedback condition and to provide positive evidence for the relationship (or otherwise) between brain responses that are suppressed during speech production and those involved in "error" detection.

Two studies did include an analysis that explicitly contrasted listening, manipulated speech, and unmanipulated speech (Meekings et al., 2016; Zheng et al., 2010), and these coordinates were included in the meta-analysis. The rest of the coordinates included in the meta-analysis came from comparisons of manipulated self-perception with unmanipulated self-perception, meaning that the effect of a listening confound cannot be ruled out. The activation seen in STG was stronger in the right hemisphere, which has been implicated in perception of pitch and spectral modulation (Hyde, Peretz, & Zatorre, 2008; Zatorre & Belin, 2001; Johnsrude, Penhune, & Zatorre, 2000), suggesting that it is possible that convergence between studies in the right hemisphere was driven by perception of pitch manipulation in FAF studies.

None of the studies considered here demonstrated all three components of a feedback response, that is, speaking-induced suppression, and an STG response to altered feedback that correlates with behavioral compensation. The only study not to use a paradigm that altered the sensory consequences of speech production, instead looking to overtly induce speech errors (Abel et al., 2009), did not find any STG activation. However, the errors reported were mainly semantic rather than phonetic slips, so it is possible that talkers in this experiment were recruiting a different type of error detection and correction network. In addition, the other studies were looking at short-term compensation processes to perturbed feedback, in utterances that were in many cases deliberately slowed to allow time for a compensatory response. Because most speech errors happen too quickly to evoke compensation, it is more likely that the neural activation found by Abel and colleagues reflects longer-term adaptation processes, whereby mistakes cause

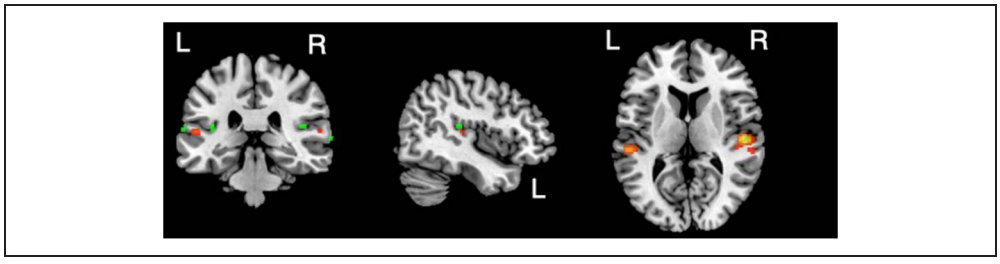

Figure 9. Regions of significant convergence between altered feedback studies (red/orange) shown at coordinates [44, −30.7, 15.1] in MNI space. The activation borders on auditory error/target map coordinates proposed by the DIVA model (green) in the right hemisphere, but there is no overlap between the two.

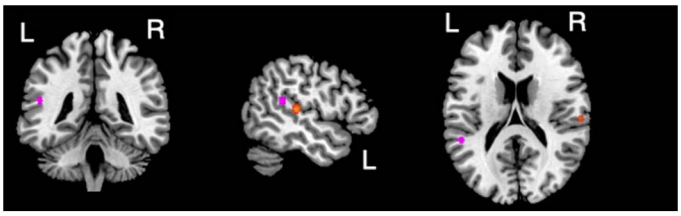

Figure 10. Regions of significant convergence between altered feedback studies (red/orange) shown at coordinates [−51, −42, 18] in MNI space. Activation is anterior to the mean location of region Spt (pink) as described by Hickok et al. (2009), and there is no overlap between the two.

internal representations to become updated. Although current models do not differentiate between the neural systems involved in compensation versus adaptation, it is possible that adaptation recruits different networks, resulting in the activation pattern seen here.

DISCUSSION

Current fMRI and PET research offers insufficient evidence to support the conclusion that STG functions as an error monitor during speech production. We reach this conclusion for two distinct reasons, which we frame below as "problems," with suggestions for how each could be addressed.

The first problem is that we found little consistency between dorsolateral temporal and inferior parietal fields that have been argued to be important for error detection and the results of our meta-analysis. Thus, although areas of significant convergence between studies revealed by our meta-analysis did border on one region in the right planum temporale identified as a putative error correction region by the DIVA model, there was no actual overlap between this proposed error correction region and the ALE map. There was also no overlap between the ALE clusters we have identified and four other ROIs defined either as auditory error correction regions by the DIVA model or as the site of auditory–motor transform in the HSFC model. Our method of comparison was relatively crude, intending only to establish where the proposed DIVA and HSFC model components fell in relation to the aggregated fMRI and PET results. We expected the ROIs based on model coordinates to fall comfortably within the ALE map, as the component locations have been proposed partly based on the existing literature, although many of the studies considered in this paper framed their results as supporting the DIVA or HSFC model. The lack of consistency between the results of our meta-analysis and the predictions from the published papers is surprising and perhaps reflects the restrictions of our comparison. It is to be expected that the activation in each model component spreads beyond the hypothesized peaks, which we attempted to simulate by creating 3-mm ROIs rather than focusing on specific voxels. The best comparison, for which we lacked the resources, would be to run a full experimental simulation with the DIVA model and compare the modeled activation with the ALE results. In this case, it is probable that there would be some overlap in activation, for example, in the right posterior STG where the ALE map bordered on one of our DIVA ROIs. It is also the case that the anatomical and functional heterogeneity of STG means that one would expect anterior and posterior fields to be differentially recruited in perceptual processing (Jasmin, Lima, & Scott, 2019). For example, Osnes and colleagues (Osnes, Hugdahl, Hjelmervik, & Specht, 2011) found higher sensitivity for sounds containing phonetic (vs. nonphonetic) features in middle STG and anterior planum temporale bilaterally. This is similar to the pattern of activation found in our ALE meta-analysis, perhaps reflecting the fact that manipulating self-perception of speech through DAF or FAF necessarily involves playing sounds with phonetic content to the talker.

The mismatch between the neural systems recruited by "error detection" and those identified in our ALE analysis may also partly arise from our second problem, which is that many of these studies are methodologically and theoretically inconsistent with one another.

From a systematic assessment of the evidence, it seems probable that many of the studies designed to test speech error correction mechanisms may not have all been testing the same behavior. Whereas models are explicit in identifying that auditory targets are defined at the syllable level, the exact level of the "error" induced by feedback manipulation varies across experiment types. Thus, Abel et al.'s (2009) picture-naming experiment elicited semantic errors. In the case of FAF, the "error" may be utterance-level pitch or formant frequency, depending on the type of manipulation. Presumably, the "error" when participants speak in noise is utterance-level intensity, pitch, and spectral center of gravity, because these are the acoustic features that participants most often change in compensation (Cooke & Lu, 2010). For DAF, the error may be the coordination of somatosensory with auditory feedback. These not only are all different kinds of "error" but also seem different from Hickok's and Guenther's definition of an auditory target (i.e., a syllable), and yet all three types of study have been cited as evidence in support of each model's conceptualization of the feedback loop and error correction mechanisms. There are very few features that are common to all of the experiments reported here. Even studies that used the same altered feedback

technique rarely used the same type or degree of manipulation. This means that, to argue that all studies are probing the same mechanism, we would need solid behavioral evidence that they prompt similar behavior (i.e., "error correction" or compensating for the altered feedback), correlating with any activation in "error correction" areas of the brain. This evidence is somewhat lacking: Many authors did not look for behavioral compensation, whereas those who did found variable results that did not always correlate with neural activation.

Although beyond the scope of this review, evidence regarding the neural underpinnings of speech feedback control is not limited to fMRI and PET. Studies using electrophysiological techniques (EEG and MEG) have demonstrated modulation of cortical responses to altered and unaltered feedback, contrasted with listening. Although the spatial resolution of these techniques is limited, they offer a complementary perspective on speech motor control by providing a snapshot of the temporal dynamics of the neural response to altered feedback. These studies focus on the ERP component N100 (sometimes known as N1) and the M100, its MEG equivalent. This is a large negative ERP, peaking 80–120 msec after stimulus, that typically occurs in response to auditory input or any unpredictable stimulus (Du et al., 2017; Sur & Sinha, 2009). The signal is distributed over fronto-central regions in EEG and can be narrowed to the supratemporal source in MEG. The N100/M100 has been shown to exhibit speaking-induced suppression in EEG (Behroozmand & Larson, 2011; Ford et al., 2001) and MEG (Houde, Nagarajan, Sekihara, & Merzenich, 2002) compared to listening, with an increase in the response when speaking with pitch-shifted feedback (Heinks-Maldonado, Nagarajan, & Houde, 2006) or in masking noise (Houde et al., 2002). This supports the basic underlying concept of feedback sensitivity in STG (although similar concerns about the naturalness of speech production tasks, and evidence of behavioral compensation, apply). However, if we want to enhance our understanding of the specific anatomical basis of speech motor control, as may be important for understanding neurological speech disorders, we need to revisit standard ways of testing this in fMRI.

Conclusions

The 17 studies discussed here presented activation in various areas of STG as support for the idea that specific areas of posterior STG are implicated in error monitoring and feedback control. We assessed each study in a systematic review and found that the quality of experiments varied and, in many cases, was not sufficient to draw strong conclusions about the role of posterior STG in auditory feedback control and error correction. The speech production tasks used were frequently highly constrained rather than naturalistic, whereas the type of "error" evoked by each perturbation is not always consistent with the syllable-level auditory targets specified by neural models of speech production. In addition, a lack of behavioral data demonstrating that adaptation or compensation is actually occurring limits the conclusions that can be drawn from neural data. Similarly, many studies did not control for the auditory effects of the perturbation by including a listening condition, meaning that they also could not establish the presence of speaking-induced suppression, a key part of the hypothesized auditory feedback loop. Because of these methodological issues, the interpretation of the ALE analysis is similarly constrained. We cannot definitively say that either the existing literature or the ALE meta-analysis we have performed provides evidence for error monitoring or auditory feedback control in posterior STG.

However, we did identify bilateral auditory fields associated with changes in the perceptual consequences of speaking aloud: These lay in the anterior auditory cortex. Further studies will be able to determine the functional role(s) of these auditory cortical fields in perceptual processing and the links to speech production.

It is important that future research include both a listening control condition and some measures of behavioral compensation to perturbation, so that confident conclusions about the role of posterior STG in error monitoring and speech production may be drawn. We also argue that experiments should move toward the use of more naturalistic speech production tasks that specifically target the type of "error" found in daily life and how manipulations are affecting self-perception. A recent fMRI study by Gauvin et al. (2016) successfully used tongue twisters to elicit phonemic errors in an investigation of prearticulatory speech monitoring; if this paradigm can be applied to post-articulatory monitoring, this is likely to be the best possible way to evaluate the neural basis of error monitoring and correction processes in speech production. Importantly, the Gauvin et al. study uses a paradigm where talkers are not only highly aware of errors but also aware that their speech production is difficult and that errors are likely. Whatever speech tasks we use in the scanner, we need to consider both the demands of that speech production task and the participants' experience of what their communicative goals might be, accordingly.

Acknowledgments

This research was supported by an Economic and Social Research Council Studentship granted to S. M.

Reprint requests should be sent to Sophie Meekings, Percy Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom, or via e-mail: Sophie.meekings@ncl.ac.uk.

REFERENCES

Abel, S., Dressel, K., Kümmerer, D., Saur, D., Mader, I., Weiller, C., et al. (2009). Correct and erroneous picture naming responses in healthy subjects. *Neuroscience Letters*, *463*, 167–171. **DOI:** https://doi.org/10.1016/j.neulet.2009.07.077, **PMID:** 19647038

Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., et al. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping*, *26*, 44–53. **DOI:** https://doi.org/10.1002/hbm.20120, **PMID:** 15884023, **PMCID:** PMC6871759

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*, 54. **DOI:** https://doi.org/10.1186/1471-2202-12-54, **PMID:** 21645406, **PMCID:** PMC3120724

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., et al. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *Neuroimage*, *109*, 418–428. **DOI:** https://doi.org/10.1016/j.neuroimage.2015.01.040, **PMID:** 25623499, **PMCID:** PMC4339397

Bennett, C. M., Baird, A. A., Miller, M. B., & Wolford, G. L. (2004). Neural correlates of interspecies perspective taking the post mortem atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, *1*, 1–5. **DOI:** https://doi.org/10.1016/S10538119(09)712029

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. S. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, *125*, 1829–1838. **DOI:** https://doi.org/10.1093/brain/awf191, **PMID:** 12135973

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage*, *16*, 497. **DOI:** https://doi.org/10.1016/S1053-8119(02)90010-8

Brett, M., Christoff, K., Cusack, R., & Lancaster, J. (2001). Using the Talairach atlas with the MNI template. *Neuroimage*, *13*, S85. **DOI:** https://doi.org/10.1016/S1053-8119(01)91428-4

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*, 687–697. **DOI:** https://doi.org/10.1016/J.NEURON.2005.09.029, **PMID:** 16301183

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*, 3153–3161. **DOI:** https://doi.org/10.1121/1.423073, **PMID:** 9637026

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, *31*, 1327–1342. **DOI:** https://doi.org/10.1016/j.neuroimage.2006.01.036, **PMID:** 16546406

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 2653–2658. **DOI:** https://doi.org/10.1073/pnas.1216827110, **PMID:** 23345447, **PMCID:** PMC3574939

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*, 868–879. **DOI:** https://doi.org/10.1002/hbm.20315, **PMID:** 17266104, **PMCID:** PMC6871445

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*, e18307. **DOI:** https://doi.org/10.1371/journal.pone.0018307, **PMID:** 21625532, **PMCID:** PMC3098236

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America*, *128*, 2059–2069. **DOI:** https://doi.org/10.1121/1.3478775, **PMID:** 20968376

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 13494–13499. **DOI:** https://doi.org/10.1073/pnas.93.24.13494, **PMID:** 8942962, **PMCID:** PMC33636

Du, X., Choa, F.-S., Summerfelt, A., Rowland, L. M., Chiappelli, J., Kochunov, P., et al. (2017). N100 as a generic cortical electrophysiological marker based on decomposition of TMS-evoked potentials across five anatomic locations. *Experimental Brain Research*, *235*, 69–81. **DOI:** https://doi.org/10.1007/s00221-016-4773-7, **PMID:** 27628235, **PMCID:** PMC5269602

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

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *Neuroimage*, *59*, 2349–2361. **DOI:** https://doi.org/10.1016/j.neuroimage.2011.09.017, **PMID:** 21963913, **PMCID:** PMC3254820

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*, 2907–2926. **DOI:** https://doi.org/10.1002/hbm.20718, **PMID:** 19172646, **PMCID:** PMC2872071

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, U.S.A.*, *113*, 7900–7905. **DOI:** https://doi.org/10.1073/pnas.1602413113, **PMID:** 27357684, **PMCID:** PMC4948312

Eliades, S. J., & Wang, X. (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex*, *15*, 1510–1523. **DOI:** https://doi.org/10.1093/cercor/bhi030, **PMID:** 15689521

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, *453*, 1102–1106. **DOI:** https://doi.org/10.1038/nature06910, **PMID:** 18454135

Eliades, S. J., & Wang, X. (2012). Neural correlates of the lombard effect in primate auditory cortex. *Journal of Neuroscience*, *32*, 10737–10748. **DOI:** https://doi.org/10.1523/JNEUROSCI.3448-11.2012, **PMID:** 22855821, **PMCID:** PMC3752069

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America*, *70*, 45–50. **DOI:** https://doi.org/10.1121/1.386580, **PMID:** 7264071

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. *Journal of Neuroscience*, *30*, 16643–16650. **DOI:** https://doi.org/10.1523/JNEUROSCI.1809-10.2010, **PMID:** 21148003, **PMCID:** PMC3010242

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry*, *158*, 2069–2071. **DOI:** https://doi.org/10.1176/appi.ajp.158.12.2069, **PMID:** 11729029

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., et al. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*, 969–977. **DOI:** https://doi.org/10.1093/cercor/bhj039, **PMID:** 16195470

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *Neuroimage*, *126*, 96–105. **DOI:** https://doi.org/10.1016/j.neuroimage.2015.11.037, **PMID:** 26608243

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*, *52*, 862–874. **DOI:** https://doi.org/10.1016/J.NEUROIMAGE.2009.10.023, **PMID:** 19837177, **PMCID:** PMC2891349

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). FMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage*, *55*, 1324–1338. **DOI:** https://doi.org/10.1016/j.neuroimage.2010.12.065, **PMID:** 21195191, **PMCID:** PMC3065208

Gracco, V. L., Tremblay, P., & Pike, B. (2005). Imaging speech production using fMRI. *Neuroimage*, *26*, 294–301. **DOI:** https://doi.org/10.1016/j.neuroimage.2005.01.033, **PMID:** 15862230

Greenlee, J. D. W., Behroozmand, R., Larson, C. R., Jackson, A. W., Chen, F., Hansen, D. R., et al. (2013). Sensory–motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS One*, *8*, e60783. **DOI:** https://doi.org/10.1371/journal.pone.0060783, **PMID:** 23577157, **PMCID:** PMC3620048

Guenther, F. H., & Hickok, G. (2015). Role of the auditory system in speech production. *Handbook of Clinical Neurology*, *129*, 161–175. **DOI:** https://doi.org/10.1016/B978-0-444-62630-1.00009-3, **PMID:** 25726268

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *Journal of the Acoustical Society of America*, *123*, 273–278. **DOI:** https://doi.org/10.1121/1.2817357, **PMID:** 18177157

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*, 22–28. **DOI:** https://doi.org/10.1002/hbm.10119, **PMID:** 12953303, **PMCID:** PMC6871912

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport*, *17*, 1375–1379. **DOI:** https://doi.org/10.1097/01.WNR.0000233102.43526.E9, **PMID:** 16932142, **PMCID:** PMC4060597

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*, 263–270. **DOI:** https://doi.org/10.1016/S0028-3932(02)00159-8, **PMID:** 12457752

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, *29*, 2–20. **DOI:** https://doi.org/10.1080/01690965.2013.834370, **PMID:** 24489420, **PMCID:** PMC3904400

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory–motor integration for speech processing. *Journal of Neurophysiology*, *101*, 2725–2732. **DOI:** https://doi.org/10.1152/jn.91099.2008, **PMID:** 19225172

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., et al. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, *8*, 2379–2382. **DOI:** https://doi.org/10.1097/00001756-199707070-00055, **PMID:** 9243644

Holmes, C., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, *22*, 324–333. **DOI:** https://doi.org/10.1097/00004728-199803000-00032, **PMID:** 9530404

Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, *33*, 174–181. **DOI:** https://doi.org/10.1016/j.conb.2015.04.006, **PMID:** 25989242, **PMCID:** PMC4628828

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82. **DOI:** https://doi.org/10.3389/fnhum.2011.00082, **PMID:** 22046152, **PMCID:** PMC3200525

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*, 1125–1138. **DOI:** https://doi.org/10.1162/089892902760807140, **PMID:** 12495520

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*, 632–639. **DOI:** https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.09.004, **PMID:** 17959204

Jasmin, K., Lima, C. F., & Scott, S. K. (2019). Understanding rostral–caudal auditory cortex contributions to auditory perception. *Nature Reviews Neuroscience*, *20*, 425–434. **DOI:** https://doi.org/10.1038/s41583-019-0160-2, **PMID:** 30918365, **PMCID:** PMC6589138

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*, 155–163. **DOI:** https://doi.org/10.1093/brain/123.1.155, **PMID:** 10611129

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, 55, 295–308. **DOI:** https://doi.org/10.1177/0023830911417802, **PMID:** 22783636

Kell, C. A., Darquea, M., Behrens, M., Cordani, L., Keller, C., & Fuchs, S. (2017). Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading. *Human Brain Mapping*, *38*, 493–508. **DOI:** https://doi.org/10.1002/hbm.23398, **PMID:** 27622923, **PMCID:** PMC6866884

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*, 535–540. **DOI:** https://doi.org/10.1038/nn.2303, **PMID:** 19396166, **PMCID:** PMC2841687

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*, 9351–9358. **DOI:** https://doi.org/10.1523/JNEUROSCI.0404-12.2012, **PMID:** 22764242, **PMCID:** PMC3404292

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*, 677–710. **DOI:** https://doi.org/10.1044/jshr.1404.677

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

Meekings, S., Boebinger, D., Evans, S., Lima, C. F., Chen, S., Ostarek, M., et al. (2015). Do we know what we're saying? The roles of attention and sensory information during speech production. *Psychological Science*, *26*, 1975–1977. **DOI:** https://doi.org/10.1177/0956797614563766, **PMID:** 26464309, **PMCID:** PMC4871256

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., et al. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *Journal of the Acoustical Society of America*, *140*, 8–19. **DOI:** https://doi.org/10.1121/1.4948587, **PMID:** 27475128

Mueller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research*, *215*, 61–76. **DOI:** https://doi.org/10.1016/0006-8993(81)90491-1

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

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, *33*,

12090–12098. **DOI:** https://doi.org/10.1523/JNEUROSCI.1008-13.2013, **PMID:** 23864694, **PMCID:** PMC3713738
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). New York: Academic Press.
Osnes, B., Hugdahl, K., Hjelmervik, H., & Specht, K. (2011). Increased activation in superior temporal gyri as a function of increment in phonetic features. *Brain and Language*, *16*, 97–101. **DOI:** https://doi.org/10.1016/j.bandl.2010.10.001, **PMID:** 21055799
Osser, H., & Peng, F. (1964). A cross cultural study of speech rate. *Language and Speech*, *7*, 120–125. **DOI:** https://doi.org/10.1177/002383096400700208
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*, 314–322. **DOI:** https://doi.org/10.1016/j.neuroimage.2012.02.068, **PMID:** 22406500, **PMCID:** PMC3342468
Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J.-L., Nazarian, B., et al. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*, *50*, 626–638. **DOI:** https://doi.org/10.1016/J.NEUROIMAGE.2009.12.065, **PMID:** 20035884, **PMCID:** PMC2840383
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*, 713–723. **DOI:** https://doi.org/10.1006/nimg.2002.1105, **PMID:** 12169255
Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acustica*, *86*, 1038–1045.
Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*, 97–132. **DOI:** https://doi.org/10.1016/S0010-0277(00)00090-1
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724. **DOI:** https://doi.org/10.1038/nn.2331, **PMID:** 19471271, **PMCID:** PMC2846110
Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition*, *68*, 204–214. **DOI:** https://doi.org/10.1016/J.BANDC.2008.04.007, **PMID:** 18541355
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. **DOI:** https://doi.org/10.1037/0033-2909.86.3.638
Rousey, C., & Holzman, P. S. (1967). Recognition of one's own voice. *Journal of Personality and Social Psychology*, *6*, 464–466. **DOI:** https://doi.org/10.1037/h0024837, **PMID:** 6082480
Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration*, *14*, 99–106. **DOI:** https://pdfs.semanticscholar.org/ad26/5c494be00224f1d10a14684b2dce331efc01.pdf
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*, 852–861. **DOI:** https://doi.org/10.1037/0278-7393.16.5.852
Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. **DOI:** https://doi.org/10.1136/BMJ.D4002, **PMID:** 21784880
Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*, 1004–1006. **DOI:** https://doi.org/10.1038/nn.2163, **PMID:** 19160497, **PMCID:** PMC2747248
Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*, 70–73. **DOI:** https://doi.org/10.4103/0972-6748.57865, **PMID:** 21234168, **PMCID:** PMC3016705
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*, *53*, 226–236. **DOI:** https://doi.org/10.1044/1092-4388(2009/09-0009)
Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, *58*, 894–901. **DOI:** https://doi.org/10.1016/j.jclinepi.2005.01.006, **PMID:** 16085192
Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*, 952–981. **DOI:** https://doi.org/10.1080/01690960903498424, **PMID:** 23667281, **PMCID:** PMC3650855
Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*, 1429–1443. **DOI:** https://doi.org/10.1016/j.neuroimage.2007.09.054, **PMID:** 18035557, **PMCID:** PMC3658624
Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., et al. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*, 499–503. **DOI:** https://doi.org/10.1016/j.neuroscience.2007.02.023, **PMID:** 17395381
Troiani, V., Fernández-Seara, M. A., Wang, Z., Detre, J. A., Ash, S., & Grossman, M. (2008). Narrative speech production: An fMRI study using continuous arterial spin labeling. *Neuroimage*, *40*, 932–939. **DOI:** https://doi.org/10.1016/J.NEUROIMAGE.2007.12.002, **PMID:** 18201906, **PMCID:** PMC2291537
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*, 1–13. **DOI:** https://doi.org/10.1002/hbm.21186, **PMID:** 21305667, **PMCID:** PMC4791073
Viechtbauer, W. (2015). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v036.i03
Wise, R. J., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, *353*, 1057–1061. **DOI:** https://doi.org/10.1016/s0140-6736(98)07491-1, **PMID:** 10199354
Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48*, 607–618. **DOI:** https://doi.org/10.1016/j.neuropsychologia.2009.10.025, **PMID:** 19896958
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953. **DOI:** https://doi.org/10.1093/cercor/11.10.946, **PMID:** 11549617
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*, 1770–1781. **DOI:** https://doi.org/10.1162/jocn.2009.21324, **PMID:** 19642886, **PMCID:** PMC2862116
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. *Journal of Neuroscience*, *33*, 4339–4348. **DOI:** https://doi.org/10.1523/JNEUROSCI.6319-11.2013, **PMID:** 23467350, **PMCID:** PMC3673229