Neural Mechanisms for Monitoring and Halting of Spoken Word Production

Samuel J. Hansen[1], Katie L. McMahon[2], and Greig I. de Zubicaray[2]

Abstract

■ During conversation, speakers monitor their own and others' output so they can alter their production adaptively, including halting it if needed. We investigated the neural mechanisms of monitoring and halting in spoken word production by employing a modified stop signal task during fMRI. Healthy participants named target pictures and withheld their naming response when presented with infrequent auditory words as stop signals. We also investigated whether the speech comprehension system monitors inner (i.e., prearticulatory) speech via the output of phonological word form encoding as proposed by the perceptual loop theory [Levelt, W. J. M. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press, 1989] by presenting stop signals phonologically similar to the target picture name (e.g., cabbage–CAMEL). The contrast of successful halting versus naming revealed extensive BOLD signal responses in bilateral inferior frontal gyrus, preSMA, and superior temporal gyrus. Successful versus unsuccessful halting of speech was associated with increased BOLD signal bilaterally in the posterior middle temporal, frontal, and parietal lobes and decreases bilaterally in the posterior and left anterior superior temporal gyrus and right inferior frontal gyrus. These results show, for the first time, the neural mechanisms engaged during both monitoring and interrupting speech production. However, we failed to observe any differential effects of phonological similarity in either the behavioral or neural data, indicating monitoring of inner versus external speech might involve different mechanisms. ■

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

INTRODUCTION

Speaking is often described as fluid and effortless (Levelt, 1983, 1989). However, speakers do need to monitor their production to ensure it is contextually appropriate and sometimes halt it altogether when either an internal error is detected or when an interlocutor attempts to interrupt a conversation (Levelt, 1983; Postma, 2000). Impairments of speech monitoring and control feature in a number of neuropsychological disorders such as Tourette's syndrome, schizophrenia, and stuttering. Tourette's syndrome is typically characterized by chronic, involuntary vocal tics. Coprolalia, the unprovoked production of socially inappropriate remarks, is also present in a subset of patients (Shapiro, Shapiro, Young, & Feinberg, 1988). The auditory hallucinations experienced by people with schizophrenia have been proposed to reflect an impairment in speech monitoring (David, 1999; McGuire et al., 1995). Stuttered speech has been linked to impairments in both self-monitoring and motor control processes (Loucks, Kraft, Choo, Sharma, & Ambrose, 2011; Brown, Ingham, Ingham, Laird, & Fox, 2005; Fox et al., 1996). Postma and Kolk (1993) propose that an abnormally slow rate of phonological encoding results in the incorporation of a large number of phonological errors into speech plans of intended utterances and that attempts to covertly repair these errors underlie the symptoms of stuttering.

At the neural level, stuttering has been broadly characterized by "underactivity" lateralized to the language-dominant left hemisphere (more specifically, the auditory areas implicated in self-monitoring) and "overactivity" largely lateralized to right hemisphere motor areas (Loucks et al., 2011; Fox et al., 1996). The right inferior frontal gyrus (IFG) has also consistently been implicated in stuttering, revealing differential activation compared with normal speaking participants for both internal (e.g., phoneme monitoring) and external (e.g., picture naming) speech processing tasks (Loucks et al., 2011). The functional role of the right IFG in stuttering has been interpreted as reflecting atypical speech planning processes (Lu et al., 2010) and/or processes that may actually compensate for the deficits typically associated with stuttering (see review by Etchell, Civier, Ballard, & Sowman, 2018). However, we know relatively little about the neural mechanisms responsible for monitoring and halting speech in healthy speakers.

According to Levelt's (1983, 1989) comprehensive and influential "perceptual loop" theory, the same speech comprehension system used to parse and understand language spoken by others monitors self-produced speech via two processing loops. The inner loop uses prearticulatory speech as input, that is, the phonologically encoded phonemic plan to be sent to the motor articulators, with

[1]University of Queensland, [2]Queensland University of Technology

© 2019 Massachusetts Institute of Technology

word-initial segments monitored more quickly than word final ones (Wheeldon & Levelt, 1995). The external loop uses overtly articulated speech as input, that is, physical soundwaves via the ears, much like listening to others speaking. Levelt (1983) also proposed a rule that governs halting speech known as the "main interruption" rule. According to this rule, upon detection of an internal error, a speaker will attempt to interrupt themselves as soon as possible. Computational models estimate the process to take approximately 150 msec (±50 msec) from detection of a speech error to interruption of the speech flow (Hartsuiker & Kolk, 2001). Empirical studies also corroborate this estimate (Ladefoged, Silverstein, & Papcun, 1973) with comparable interruption times for other nonspeech actions, for example, typing (Logan & Cowan, 1984). Indefrey's (2011) meta-analysis (see also Indefrey & Levelt, 2004) of the cerebral correlates of spoken word production attributed self-monitoring to the left superior temporal gyrus (STG). However, the meta-analysis was not able to differentiate regions of the STG engaged by inner versus outer speech monitoring loops and did not address potential mechanisms engaged for halting speech. More recent neuroimaging research has challenged a role for the STG in internal speech monitoring. For example, Gauvin, De Baene, Brass, and Hartsuiker (2016) were unable to detect significant activation in the STG during speech production when auditory feedback from self-produced speech was masked by noise.

One paradigm that has been employed frequently to assess withholding or halting of responses outside the language domain is the stop signal paradigm (Logan, 1981). The classic version of the paradigm requires participants to respond as quickly as possible via manual button press to a pair of target "go" stimuli presented on multiple trials (e.g., the letters X and O, or geometric shapes such as a triangle and circle). On a subset (e.g., 25%) of trials, a stop signal (e.g., auditory tone) is presented after a brief variable delay following the target indicating participants are to withhold their response on that trial. If the participant successfully stops, an adaptive calibrating algorithm increases the delay between go and stop signal presentation (known as the stop signal delay) by 50 msec, making halting on the next stop signal trial slightly more difficult. Conversely, if the participant was unsuccessful at stopping, the delay for the subsequent stop signal decreases by 50 msec, making halting on the next stop signal trial slightly easier. This one-up/one-down stepping algorithm ensures approximately 50% halting accuracy in the stop signal task. By subtracting the average stop signal delay from the average go RT, a stop signal RT (SSRT) can be calculated that provides a measure of how quickly a participant can withhold their response. According to a horse race model of dual task processing, independent go and stop processes race against each other to completion to determine if a response is initiated or inhibited (Verbruggen & Logan, 2009). Neuroimaging studies have attributed stopping performance to a domain-general, inhibitory control mechanism mediated by a mostly right-lateralized circuit involving the IFG and preSMA (e.g., Aron, Robbins, & Poldrack, 2014).

To our knowledge, only one neuroimaging study has investigated the neural mechanisms engaged specifically for halting speech in response to an external interruption cue. Xue, Aron, and Poldrack (2008) used fMRI and a modified stop signal paradigm in which participants read aloud or responded manually to two letters ("T" and "D") or read aloud a series of nonwords. Across all three tasks, they were instructed to withhold their response when an auditory stop signal (a tone) was presented. A conjunction analysis of "stop > go" trials across all three tasks revealed activation in the right IFG and preSMA that the authors attributed to a neural mechanism for domain-general response inhibition. Surprisingly, activation was observed in the bilateral STG only for the manual task. The authors proposed the failure to observe differential STG activation in the vocal tasks might have been due to a subtraction of the auditory stop signal and the participants' own vocal response on go trials. However, they did not compare activity during unsuccessful versus successful halting performance that might have revealed speech error monitoring activity.

In this study, we investigated the neural mechanisms for both monitoring and halting spoken word production. Although the neural mechanisms of monitoring and halting speech have been investigated separately, both are required during conversation particularly when interlocutors attempt to interrupt the speech production of their conversational partners. This latter scenario has been omitted from recent reviews of speech monitoring (e.g., Nozari & Novick, 2017). As Postma (2000) pointed out, we hear and parse other utterances to realize that the interlocutor is signaling some misunderstanding or error in our own conversation, and this entails a monitoring loop. Furthermore, we sought to differentiate the neural mechanisms responsible for monitoring prearticulatory versus overtly produced speech. To accomplish this, we employed a modified stop signal paradigm during fMRI in which participants were instructed to name target pictures aloud as quickly as possible but withhold their naming response when an auditory word was presented subsequently. Contrasting unsuccessful halting (failed stop) versus successful halting should reveal the brain regions involved in speech error commission and monitoring in relation to an external cue interrupting speech production. Based on prior work on speech monitoring and domain-general control mechanisms, we predicted involvement of STG (e.g., Indefrey, 2011), IFG, preSMA, and ACC, respectively—the latter regions classically associated with response inhibition in the stop signal task for the contrast of successful halting versus response initiation (Piai, Roelofs, Acheson, & Takashima, 2013; Indefrey, 2011; Xue et al., 2008; Christoffels, Formisano, & Schiller, 2007). However, given the stop condition

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

involves hearing an auditory word and the go condition involves hearing self-produced speech, we do not expect to observe STG activation for contrasts of go versus successful stop conditions, consistent with Xue et al.'s (2008) results.

To differentiate the inner and external speech monitoring mechanisms proposed by Levelt's (1983, 1989) perceptual loop account, we used words as stop signals that were either phonologically similar (i.e., shared an initial phoneme) or dissimilar to the target picture names (e.g., cabbage–CAMEL vs. button–CAMEL). Prior work indicates speakers are able to monitor their internal speech production at the word form level with a relative advantage for word-initial segments/phonemes (Nooteboom & Quené, 2019; Wheeldon & Levelt, 1995). There is clear evidence in the working memory and attention literatures of a privileged connection between hearing a spoken word and comprehending it, such that avoiding comprehension is unlikely (McLeod & Posner, 1984; Salamé & Baddeley, 1982). According to the perceptual loop account, when stop signals have initial phonemes that overlap with the target name, this should result in slower and less accurate halting performance compared with unrelated stop signals because the comprehension system's comparison of the inner and external speech representations will not reveal they are inconsistent until the end of the word. By contrast, unrelated stop signals are discriminated earlier by the comprehension system as their initial phonemes do not match those of the intended target and so are able to be acted on more quickly. Contrasting successful halting to phonologically related versus unrelated stop signals should therefore differentiate activity in the STG (Wernicke's area) proposed to be responsible for monitoring prearticulatory, inner speech via the comprehension system (Indefrey, 2011).

METHODS

Participants

We recruited 20 healthy volunteers (14 women, six men) aged from 19 to 34 years ($M = 25.25$, $SD = 4.42$). Each participant provided written informed consent and was compensated AUD$30 for their time and effort. All participants identified as monolingual English speakers, right-handed, with normal or corrected-to-normal vision and hearing, and no history of neurological or psychiatric disorder. The study was approved by the Medical Research Ethics Committee of the University of Queensland.

Behavioral Design

This experiment used a repeated-measures design. The independent variable was the phonological relatedness of the stop signal word to the target picture name (within participant and within item) with two levels: phonologically related (e.g., BUCKET–button, CAMEL–cabbage) and unrelated (e.g., BUCKET–cabbage, CAMEL–button). The primary dependent variables were the SSRT (calculated by subtracting the mean stop signal delay from the mean naming latency for go trials) and stop signal accuracy (i.e., the percentage of successful stop signal trials). We also recorded RTs for unsuccessful stop signal trials in which the participant erroneously produced a naming response and both RT and accuracy for go trials.

Stimuli and Materials

Two black-and-white line drawings of common objects ("bucket" and "camel") were selected as to-be-named picture stimuli (Snodgrass & Vanderwart, 1980). These pictures were selected as they are semantically and phonologically unrelated, and their names begin with stop consonants producing an easily detectable hard onset. Sixty words beginning with the same initial two phonemic segments were selected as phonologically related stop signal stimuli, that is, 30 "bu-" words such as "button" and 30 "ca-" words such as "cabbage." The phonologically unrelated stop signal stimuli were created by re-pairing the word lists with their phonologically unrelated picture (see Appendix for word lists). Picture stimuli were back-projected onto a screen that the participant viewed through a mirror mounted on a head coil. Target pictures were presented centrally on a white background with the size of the pictures approximately 10 × 10 cm and subtended approximately 10° of visual angle when each participant was in position for imaging. Stop signal words were spoken by a female native Australian speaker, recorded on digital audio at a sampling frequency of 44.1 kHz, and postequalized.

Each picture was presented 240 times for a total of 480 trials. On 25% of trials (i.e., 120 trials), the auditory stop signal was presented after a variable stop signal delay. A 30-dB attenuating electrodynamic headset was used to reduce gradient noise and present auditory stimuli (MR Confon GmbH). The sound was presented using MR-compatible piezoelectric headphones (Optoacoustics Ltd.; www.optoacoustics.com). Half of the stop signal trials (i.e., 60 trials) were phonologically related, whereas the other half were phonologically unrelated. The 480 trials were separated into three blocks of 160 trials each. Stimulus presentation, response recording, and latency measurement (i.e., voice key) were accomplished via the Cogent 2000 toolbox extension (www.vislab.ucl.ac.uk/ cogent_2000.php) for MATLAB (MathWorks, Inc.). Naming responses were recorded on digital audio files using a custom-positioned fiber-optic dual-channel noise-canceling microphone (FOMR-III, Optoacoustics Ltd.; www.optoacoustics.com) attached to the head coil. Naming accuracy was verified offline using Audacity software.

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Procedure

After positioning in the MRI scanner, participants initially underwent a practice phase, consisting of 22 go trials randomly interspersed with 10 stop signal trials, followed by the task proper. Participants were instructed to name the pictures aloud as quickly as possible but withhold their naming response if they heard a word. Additionally, participants were asked to give equal importance to going and stopping and not wait for a word to appear or not before naming the picture. Participants were also asked not to speak or move during image acquisition and, in the event of a naming error, not to correct their response.

Each trial began with a black fixation point (+) presented in the center of the screen on a white background for 500 msec. This was followed by a picture presented centrally for 2000 msec or until a response was detected. The total time between trials was 2500 msec. On 25% of trials, a stop signal word was presented according to the stop signal delay. The stop signal delay for each stop trial was calculated using an adaptive calibrating/stepping algorithm. The initial delay was set at 300 msec (i.e., stop signal word presented 300 msec after picture presentation), as per previous studies (van den Wildenberg & Christoffels, 2010). The delay for subsequent stop trials either increased by 50 msec if the participant successfully stopped or decreased by 50 msec if the participant was unsuccessful at stopping, that is, produced any articulatory sound detected by the voice key.

Participants completed three blocks of 160 trials for a total of 480 trials. The order of trials within each block was pseudorandomized across participants using Mix software (van Casteren & Davis, 2006), such that each picture was presented maximally five times in a row and no more than two successive trials were from the same condition. Five different pseudorandomizations were employed, each administered to four participants. During a short break between the first and second blocks, a structural scan was acquired (see Image Acquisition section). The entire experiment took approximately 1 hr.

Image Acquisition

Images were acquired using a 3T MAGNETOM TIM Trio MRI system (Siemens Medical Solutions) with a 12-channel Matrix head coil. Functional T2*-weighted images depicting BOLD contrast were acquired using a gradient-echo EPI sequence (36 slices, repetition time = 2500 msec, echo time = 36 msec, 64 × 64 matrix, 3.3 × 3.3 mm in plane resolution, 3 mm slice thickness with 0.3 mm gap and flip angle 80°). A point spread function mapping sequence was acquired before the EPI data to correct geometric distortions (Zaitsev, Hennig, & Speck, 2004). Three series of 178 image volumes each were acquired. A 3-D T1-weighted structural image was also acquired using a magnetization-prepared rapid acquisition gradient-echo sequence (1 mm isotropic voxels). Total imaging time was approximately 45 min.

Image Analysis

Image processing and statistical analyses were performed using statistical parametric mapping software (SPM12; Wellcome Trust Centre for Neuroimaging). The first five volumes in each fMRI block were discarded. Each functional time series was first resampled using generalized interpolation to the acquisition of the middle slice in time to correct for the interleaved acquisition sequence and then motion-corrected using the INRIalign toolbox (Freire, Roche, & Mangin, 2002). A mean image was generated from the realigned series and coregistered to the T1-weighted image. The T1-weighted image was next segmented using the *Segment* routine. The DARTEL toolbox (Ashburner, 2007) was then employed to create a custom group template from the segmented gray and white matter images, and individual flow fields were used to normalize the realigned fMRI volumes to the Montreal Neurological Institute atlas T1 template. The images were resampled to 2 mm^3 voxels and smoothed with an 8-mm FWHM isotropic Gaussian kernel. Global signal effects were then estimated and removed using a voxel-level linear model (Macey, Macey, Kumar, & Harper, 2004).

We conducted a two-stage, mixed-effects model statistical analysis. Trial types corresponding to correct and erroneous responses for each stop and go condition were modeled as effects of interest, with delta functions representing each onset at individual subject level and convolved with a canonical hemodynamic response function. Low-frequency noise and signal drift were removed from the time series in each voxel with high-pass filtering (1/128 Hz). Temporal autocorrelations were estimated and removed with an autoregressive (AR1) model. Linear contrasts were applied to each participant's parameter estimates at the fixed-effects level and then entered in a group-level random effects repeated-measures ANOVA in which covariance components were estimated using a restricted maximum likelihood procedure to correct for nonsphericity (Friston et al., 2002).

As we had a priori hypotheses concerning specific neuroanatomical regions associated with various mechanisms involved in speech monitoring and control, we opted to first restrict voxel-wise analyses to a set of predefined ROIs via small volume corrections, thereby controlling for multiple comparisons only in those voxels, using labeled maximum likelihood maps from three-dimensional probabilistic atlases. We used Hammers et al.'s (2003) probabilistic atlas as it encompassed the stereotactic Montreal Neurological Institute coordinates reported. We predefined the following ROIs: right IFG and preSMA (domain-general inhibitory control: Aron et al., 2014; Xue et al., 2008), left posterior middle temporal gyrus (MTG) and STG (phonological word form encoding

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Table 1. Mean Go RTs, Stop Signal RTs, Unsuccessful Stop RTs, and Percentage of Successful Stops as a Function of Phonological Relatedness

	Total	*Phonologically Unrelated*	*Phonologically Related*
Go RT	561	—	—
Stop signal RT	255	254	257
Unsuccessful stop RT	536	537	534
Percentage of successful stops	50%	50%	50%

All RTs in milliseconds.

and lexeme retrieval: Indefrey, 2011; de Zubicaray, McMahon, Eastburn, & Wilson, 2002), left STG (self-monitoring: Indefrey, 2011), and left IFG (response-level syllabification and phonetic encoding: Indefrey, 2011). The ROI analyses were followed by an exploratory whole-brain analysis. For both analyses, we applied a height threshold of $p < .001$ and family-wise error-corrected cluster threshold of $p < .05$.

RESULTS

Behavioral Data

Data from one participant was excluded due to a technical difficulty resulting in incomplete image acquisition. Data from an additional participant was excluded due to excessive head movement during image acquisition defined as motion exceeding one voxel (3 mm) within a single imaging run. For the remaining 18 participants, nonspeech noises and technical errors (voice key malfunctions) accounted for 0.75% of the total data and were excluded from further analysis. Participant naming errors on go trials (e.g., incorrect responses, verbal disfluencies) accounted for 0.44% of the total data and were likewise excluded. RTs for go trials exceeding three standard deviations from each participant's mean were also removed, accounting for 0.71% of data. Successful halting was defined as not producing any sound, and unsuccessful halting was defined as production of any overtly articulated speech sound. Table 1 shows mean go RTs, stop signal RTs, unsuccessful stop RTs, and percentage of successful stops as a function of phonological relatedness.

Mean SSRTs, unsuccessful stop RTs, and percentage of successful stops were analyzed using repeated-measures ANOVAs with Participants as a random factor (item analyses were not conducted due to the use of only two target pictures, as per conventional stop signal studies; see van den Wildenberg & Christoffels, 2010). There were no significant differences between SSRTs, $F(1, 17) = 1.39$, $p = .254$, $\eta_p^2 = .08$, unsuccessful stop RTs, $F(1, 17) = .30$, $p = .589$, $\eta_p^2 = .02$, or the percentage of successful stops, $F(1, 17) = .06$, $p = .805$, $\eta_p^2 < .01$, for the comparison of phonologically related and unrelated conditions. Notably, the mean unsuccessful stop RT is significantly faster than the mean go RT, $t(17) = 3.85$, $p = .001$, $d = .46$, primarily because unsuccessful stops represent the lower half (i.e., faster half) of the distribution of stop trials. This also suggests that participants were not waiting for the stop signal to be presented before responding.

Imaging Data

A Priori Defined ROI Analyses

The t contrast of phonologically similar > dissimilar stop signals revealed no significant activity in any of the predefined ROIs or whole-brain analysis for either successful or unsuccessful halting, consistent with our behavioral results. No significant activity was observed in any of the ROIs for the reverse contrast either (phonologically dissimilar > similar). We therefore collapsed stop signal word types for subsequent contrasts to identify mechanisms engaged in successful versus unsuccessful halting of speech (see Table 2 and Figures 1 and 2).

The contrast of successful stop > go revealed significant activation in the left posterior and anterior STG, the middle MTG, and bilaterally in the IFG. The reverse contrast (go > successful stop) revealed significant activation in the left posterior temporal lobe. The contrast of unsuccessful stop > go trials revealed increased activity in the left posterior STG, middle MTG, and anterior temporal pole and bilaterally in the IFG. The reverse contrast (go > unsuccessful stop) revealed increased activity in the left posterior temporal lobe. Finally, for the contrast of successful > unsuccessful halting, significant activity was observed in the left posterior MTG. Activity in the left IFG ROI was on the threshold of approaching significance ($p = .05$) for this contrast ($x = -50$, $y = 44$, $z = 2$, Z score = 3.83, cluster size = 39). The reverse contrast (unsuccessful > successful stop) revealed increased activity bilaterally in the posterior STG and in the left anterior temporal pole and right IFG (see Figure 2).

Unrestricted Whole-brain Analyses

For the contrast of successful stop > go, significant activity was observed in the right STG, left posterior middle frontal gyrus (MFG), and inferior occipital gyrus (IOG). The reverse contrast (go > successful stop) revealed a large cluster of activity with a peak in the right precuneus and a second with a peak in the left ACC and extending into the SMA. The contrast of unsuccessful stop > go conditions revealed significant activity in bilateral STG and left MFG, extending into the SMA. For the reverse contrast (go > unsuccessful stop), significant activity was observed in the right precuneus and ACC (see Figure 1).

The t contrast of successful > unsuccessful stop conditions revealed significant activity in four clusters lateralized to the left hemisphere (see Figure 2) with the peak of the largest cluster in the left MTG. As Table 2 shows,

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Table 2. Cerebral Regions Showing Significant Activity for Trial Type Comparisons

	Peak				
Contrast	*x*	*y*	*z*	*Z Score*	*Cluster Size (Voxels)*
Stop > Go					
Right STG[a]	56	−24	0	Inf.	9730
Left posterior MFG[a]	−6	8	52	7.71	14706
Left IOG[a]	−36	−88	−10	4.66	602
Left posterior temporal lobe[b]	−54	−40	12	7.25	1492
Left middle MTG[b]	−64	−32	0	5.79	101
Left posterior STG[b]	−58	−24	2	7.04	1330
Left anterior STG[b]	−52	8	−4	5.63	166
Left IFG[b]	−44	10	6	6.78	1528
Right IFG[b]	42	14	6	7.83	1218
Go > Stop					
Right precuneus[a]	16	−56	20	7.17	12510
Left ACC/SMA[a]	−10	46	−2	6.36	2498
Left posterior temporal lobe[b]	−32	−40	−10	6.37	613
	−10	−58	8	5.03	127
Unsuccessful Stop > Go					
Right posterior STG[a]	56	−24	0	Inf.	10549
Left SMA/posterior MFG[a]	−6	6	58	Inf.	2414
Left posterior STG[a,b]	−62	−30	8	Inf.	10186
Left middle MTG[b]	−64	−32	0	6.17	82
Left anterior STG[b]	−52	8	−4	6.53	302
Left IFG[b]	−48	8	0	7.30	1497
Right IFG[b]	42	14	6	Inf.	1437
Go > Unsuccessful Stop					
Right precuneus[a]	14	−54	20	6.99	15459
ACC[a]	0	54	−2	5.88	3008
Left posterior temporal lobe[b]	−32	−40	10	8.80	858
Successful Stop > Unsuccessful Stop					
Left posterior MTG[a,b]	−54	−68	0	4.93	706
	−54	−46	−8	3.73	79
Left SFG[a]	−14	10	50	4.34	321
Left superior parietal lobe[a]	−6	−62	56	4.14	299
Left angular gyrus[a]	−50	−78	28	4.10	301

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Table 2. (*continued*)

Contrast	Peak			Z Score	Cluster Size (Voxels)
	x	*y*	*z*		
Unsuccessful Stop > Successful Stop					
Left posterior STG[a,b]	−62	−34	14	4.60	500
Right posterior STG[a]	62	−24	0	5.37	966
Left anterior STG[a,b]	−50	8	−16	4.16	228
Right IFG[b]	60	14	22	4.09	59

Height threshold $p < .001$, and $p < .05$ cluster family-wise error (FWE) corrected. [a] Whole-brain corrected. [b] Small volume corrected. Inf. = infinite.

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

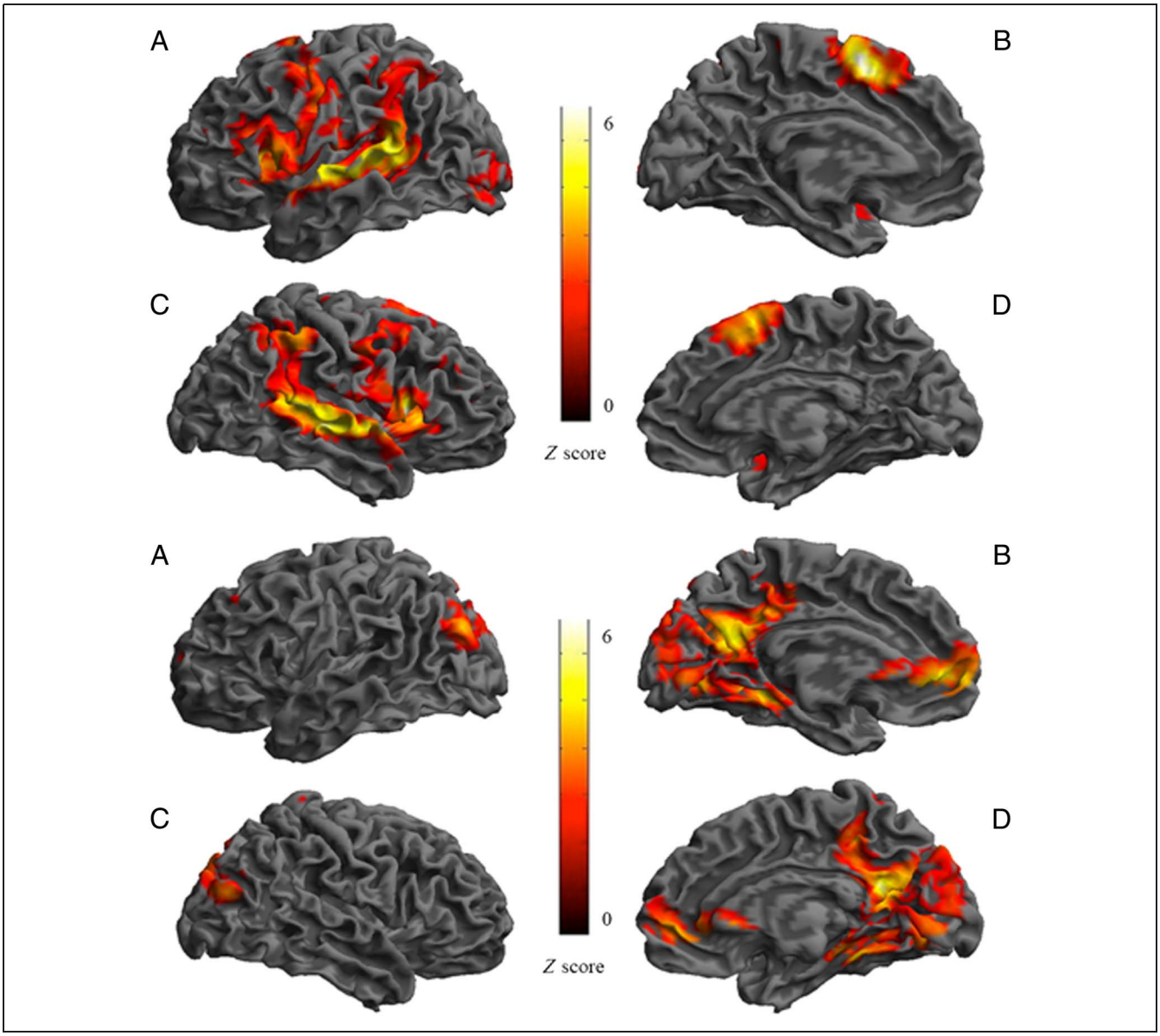

Figure 1. Cerebral regions showing significant activity for contrasts of successful stop > go (top) and go > successful stop (bottom) in the whole-brain analyses superimposed on the surface rendering of an individual brain (height threshold $p < .001$ and cluster threshold of five voxels for visualization purposes). (A) Left hemisphere lateral view, (B) left hemisphere medial view, (C) right hemisphere lateral view, and (D) right hemisphere medial view.

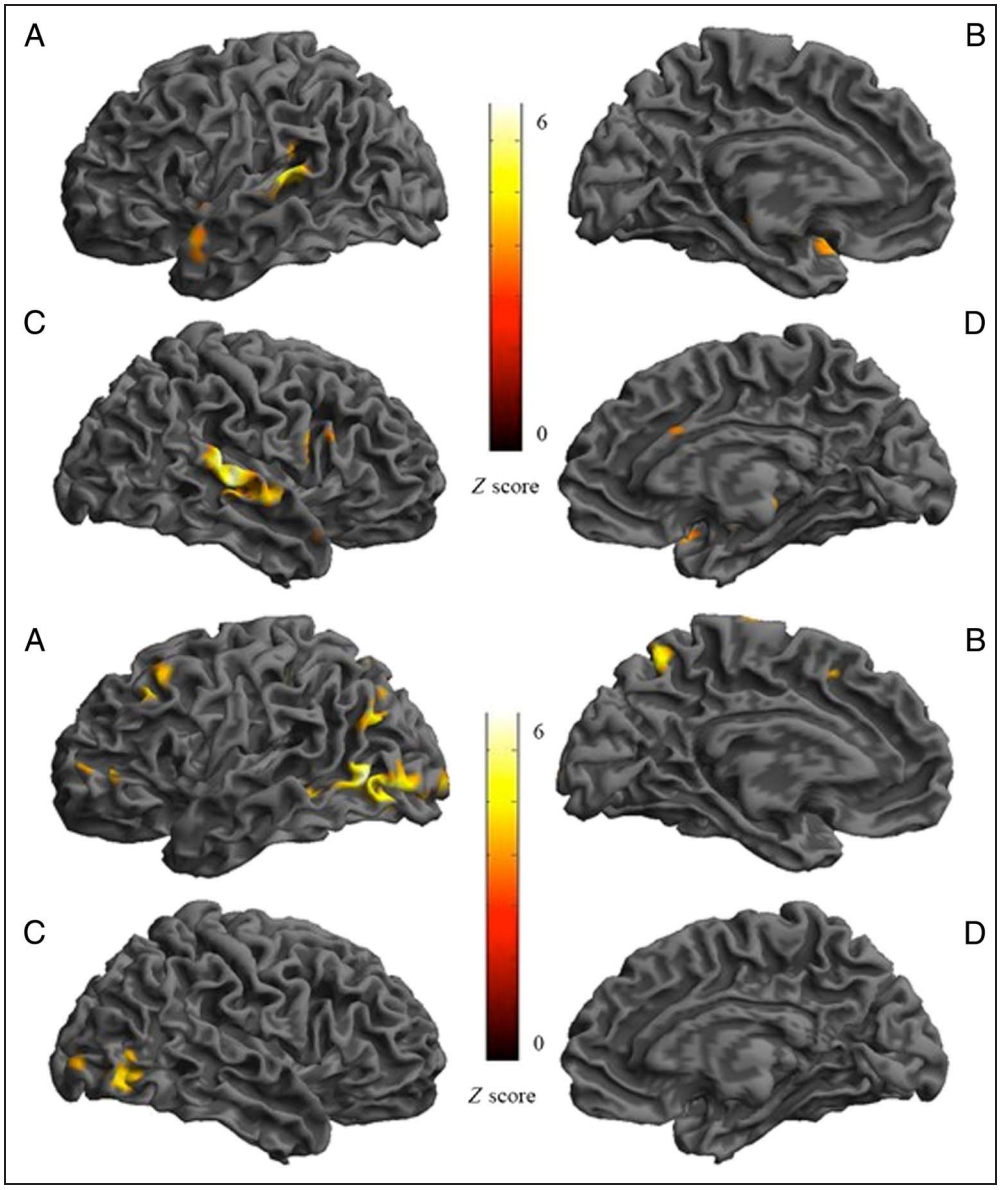

Figure 2. Cerebral regions showing significant activity for contrasts of unsuccessful > successful stop (top) and successful > unsuccessful stop (bottom) in the whole-brain analyses superimposed on the surface rendering of an individual brain (height threshold $p < .001$ and cluster threshold of five voxels for visualization purposes). (A) Left hemisphere lateral view, (B) left hemisphere medial view, (C) right hemisphere lateral view, and (D) right hemisphere medial view.

the same regions involved in the ROI analyses were also significant in the whole-brain analysis. An additional cluster on the threshold of significance ($p = .05$) was identified in the right MTG ($x = 48$, $y = -72$, $z = -8$, Z score = 4.56, cluster size = 212). For the reverse contrast (unsuccessful > successful stop conditions), significant activity was observed in three clusters localized to the bilateral temporal cortices with the peak of the largest cluster centered on the right STG.

DISCUSSION

The current study aimed to identify the neural mechanisms of monitoring and halting in speech production using a picture naming variant of the stop signal task during fMRI. Our results show, for the first time, that monitoring and interruption of speech in healthy participants is accomplished via a network of brain regions involved in spoken word production and domain-general cognitive control. However, we were not able to find any evidence for an inner speech monitoring mechanism operating at the level of phonologically encoded phonemic plans in either our behavioral or fMRI data.

Both unsuccessful versus successful halting of speech in the current paradigm entailed the same auditory input, with unsuccessful halting trials also entailing participants hearing their erroneous overt response. Overall stopping accuracy was 50%, and the magnitude of the average SSRT (259 msec) was commensurate with other studies using similar stop signal word production paradigms (van den Wildenberg & Christoffels, 2010; Slevc & Ferreira, 2006). However, it was slightly longer than Levelt's (1983, 1989) estimation of the time it takes to interrupt the speech flow according to his proposed main

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

interruption rule (see also Hartsuiker & Kolk, 2001). This might reflect the additional demands of performing the task in the MRI environment.

The fMRI results for successful stop > go trials revealed extensive activity across bilateral IFG and STG (Xue et al., 2008). The involvement of the right IFG supports proposals that this region plays a role in domain-general response inhibition during spoken word production, as it also shows significant activity in manual versions of the stop signal task (Severens, Kühn, Hartsuiker, & Brass, 2012; Xue et al., 2008). The involvement of the left STG in this contrast most likely reflects the speech comprehension system activity during detection of the stop signal word. The activation of the preSMA/SMA and ACC for the reverse contrast (go > successful stop) is consistent with the involvement of these areas in response selection/generation during spoken word production and domain-general monitoring mechanisms for potential conflict/competition (Nozari & Novick, 2017; Piai et al., 2013; de Zubicaray, Wilson, McMahon, & Muthiah, 2001). Interestingly, we observed a similar pattern of activity for the contrasts of unsuccessful stop and go trials, perhaps because the former included a large proportion of incomplete/partial responses.

We also observed significantly increased activity in bilateral posterior and left anterior STG for contrasts involving unsuccessful halting of speech and right IFG. These are novel findings because prior studies did not compare unsuccessful versus successful halting performance to reveal speech error monitoring activity (e.g., Xue et al., 2008). We interpret the STG findings as reflecting posterior auditory cortex activation based on participants' hearing their own erroneous spoken responses whereas the anterior portion likely reflects the speech-based properties of the response being produced. For example, Obleser, Wise, Dresner, and Scott (2007) reported a pattern of activation during speech comprehension with left posterior STG activating for both speech and nonspeech sounds and the anterior STG activating only for intelligible speech sounds (i.e., consonant bursts). The finding of increased right IFG activity for unsuccessful halting is interesting given the recent debate about the role of this domain-general region in posterror monitoring versus inhibitory control in stop signal paradigm performance and appears more consistent with the former interpretation (cf. Aron et al., 2014; Erika-Florence, Leech, & Hampshire, 2014; Swick & Chatham, 2014). Additionally, this finding for unsuccessful halting aligns with recent neuroimaging studies of stuttering where increased right IFG activity has been implicated in processes that may reflect attempts to compensate for the deficits typically associated with stuttering (e.g., Etchell et al., 2018).

The reverse contrast (successful > unsuccessful halting) revealed significantly increased activity bilaterally in the posterior MTG, left middle and superior frontal gyri, and the parietal lobe indicating these areas comprise an important network for successful inhibitory control of spoken responses. The posterior MTG has often been implicated in lexical-level processing, and its engagement might therefore reflect the successful interruption of lexical selection needed for halting production following detection of the stop signal (e.g., de Zubicaray & Piai, 2019; Indefrey, 2011). The engagement of a domain-general frontal-parietal network is a reliable finding for successful stop signal performance (see Erika-Florence et al., 2014). The involvement of the angular gyrus in the parietal lobe for this contrast is consistent with its established role in comprehension (Hartwigsen, Golombek, & Obleser, 2015). This might indicate that successful stopping entails attention to and successful processing of the stop signal.

According to the perceptual loop account (Nooteboom & Quené, 2019; Wheeldon & Levelt, 1995), the comprehension system uses prearticulatory, phonemically specified speech as input for monitoring of inner speech. We therefore predicted that halting performance would be significantly slower for phonologically similar stop signals, allowing us to differentiate the inner and outer monitoring loops (e.g., Slevc & Ferreira, 2006). However, neither our behavioral or fMRI data revealed any significant difference between phonologically similar and dissimilar conditions. Thus, our results are inconsistent with the proposal that inner speech monitoring operates at a phonological level of representation (Nooteboom & Quené, 2019; Slevc & Ferreira, 2006; Wheeldon & Levelt, 1995). However, they are consistent with other recent failures to support the perceptual loop account of speech monitoring. For example, Oppenheim and Dell (2008) demonstrated that, although inner speech errors do exhibit a lexical bias, they are not influenced by phonological/phonemic similarity, unlike overt errors. They surmised that inner speech must therefore be relatively impoverished at the featural level (Oppenheim & Dell, 2008).

The current study is the first to demonstrate the neural mechanisms involved in successful monitoring and halting of production. However, there are certain design limitations that should be considered when interpreting our findings. The stop signal paradigm necessarily entails "occasional" stop signals (i.e., 25% of trials), which produces an unavoidable difference in the number of trials between go and stop signal conditions. Thus, contrasts between these conditions are statistically more likely to reveal significant differences, unlike the comparisons within stop signal conditions, while also being the least theoretically interesting from the perspective of monitoring theories. In addition, we used continuous imaging rather than a sparse temporal sampling design because of the rapid naming responses (i.e., short intertrial intervals) required for the stop signal task. Thus, for the conditions involving overt production (i.e., go and unsuccessful stops), there is the possibility that speech articulation artifacts are reflected in some regions' activity, especially IFG and anterior temporal lobe where magnetic susceptibility by movement interactions inevitably occur during overt speech (e.g., Mehta, Grabowski,

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Razavi, Eaton, & Bolinger, 2006). In the contrast of successful versus unsuccessful halting, our chief comparison of interest, the largest proportion of the latter were only partial responses, so should reflect considerably less speech artifacts than the go condition.

Another potential ecological issue for the current study is that only two pictures were used as go stimuli, which is somewhat different to typical conversation with its more varied production. We opted for two go stimuli to facilitate comparisons with the conventional stop signal task (Logan, 1981) and as it allows for a more sensitive test of the effects of phonological similarity. Although item repetition in picture naming is known to speed response latencies, this effect is also known to decrease over subsequent repetitions (e.g., Griffin & Bock, 1998; Oldfield & Wingfield, 1965), and to our knowledge, no production account proposes a reduced or obviated need for the usual processes of lexical access with repetition in naming. Moreover, speech production is a highly practiced skill, usually described as effortless with speakers easily capable of producing three to four words per second with extreme accuracy (1 error in 1000; e.g., Levelt, 1989). Prior studies have shown using a series of pictures actually increases both go naming and SSRT latencies (by ~100 and ~50 msec, respectively) and their variability (van den Wildenberg & Christoffels, 2010). Hence, it is not clear that naming a set of object nouns could be considered more ecologically valid.

A final issue to address is the nature of the monitoring mechanism the stop signal task engages, as it is essentially a verbal version of a task originally designed to investigate mechanisms of response inhibition (e.g., Verbruggen & Logan, 2009). Prior work on monitoring used paradigms in which speech was not meant to be articulated, with commission errors central to interpretations about the inner loop's sensitivity to phonological representations (Nooteboom & Quené, 2019; Oppenheim & Dell, 2008; Wheeldon & Levelt, 1995). However, as we noted in the Introduction, a decision to halt speech production typically arises from two scenarios, detection of an internal error or an external cue from an interlocutor attempting to interrupt a conversational partner. Both decisions to interrupt production require comparisons with the speaker's intended speech plan, that is, they require a monitoring loop (e.g., Postma, 2000). The stop signal task is a specific example of a monitor detecting an external interruption cue, determining whether an internal speech plan is in progress, and ensuring it is halted accordingly. Thus, our findings are relevant to the latter mechanisms but are less informative for theories of monitoring for internal speech error detection and repair.

In summary, our findings show that monitoring and halting/interruption of speech in relation to an external cue is accomplished via a network of regions involved in spoken word production and domain-general cognitive control. However, we could find no evidence for an inner speech loop operating at the level of phonologically encoded phonemic plans. This was true for both our behavioral and fMRI data.

APPENDIX: WORD STIMULI LISTS

Phonologically Related to Target Picture "Bucket"			*Phonologically Related to Target Picture "Camel"*		
bud	bumper	bunkum	cab	camphor	capital
buddy	bumpkin	burrow	cabbage	campus	capstan
budget	bun	bus	cabin	camshaft	captain
buffer	bunch	bust	caboose	candid	caption
buffet	bundle	bustier	cache	candle	captive
buffoon	bungalow	butler	cad	cannon	cash
buggy	bungee	butter	caddie	canoe	cashew
bulb	bunion	button	calorie	cantor	castle
bulge	bunk	buttress	cameo	canvas	cavern
bulkhead	bunker	buzzer	camera	canyon	cavity

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Acknowledgments

We are grateful to Jennifer Burt for her helpful comments on an earlier version of this paper. The authors were supported by an Australian Postgraduate Award (S. H.), Australian Research Council Discovery Grants (DP1092619 and DP150103997 to G. Z. and K. M.) and Future Fellowship (FT0991634 to G. Z.).

Reprint requests should be sent to Samuel J. Hansen, School of Psychology, The University of Queensland, Brisbane, QLD 4072, Australia, or via e-mail: sam.hansen@uq.edu.au.

REFERENCES

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Science*, *18*, 177–185.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*, 95–113.

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*, 105–117.

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.

David, A. S. (1999). Auditory hallucinations: Phenomenology, neuropsychology and neuroimaging update. *Acta Psychiatrica Scandinavica*, *99*, 95–104.

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture–word task: An event-related fMRI study using overt vocal responding. *Neuroimage*, *16*, 1084–1093.

de Zubicaray, G. I., & Piai, V. (2019). Investigating the spatial and temporal components of speech production. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford handbook of neurolinguistics* (pp. 472–497). Oxford: Oxford University Press.

de Zubicaray, G. I., Wilson, S. J., McMahon, K. L., & Muthiah, S. (2001). The semantic interference effect in the picture-word paradigm: An event-related fMRI study employing overt responses. *Human Brain Mapping*, *14*, 218–227.

Erika-Florence, M., Leech, R., & Hampshire, A. (2014). A functional network perspective on response inhibition and attentional control. *Nature Communications*, *5*, 4073.

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*, *55*, 6–45.

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., et al. (1996). A PET study of the neural systems of stuttering. *Nature*, *382*, 158–162.

Freire, L., Roche, A., & Mangin, J.-F. (2002). What is the best similarity measure for motion correction in fMRI time series? *IEEE Transactions on Medical Imaging*, *21*, 470–484.

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage*, *16*, 484–512.

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.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313–338.

Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., et al. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, *19*, 224–247.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113–157.

Hartwigsen, G., Golombek, T., & Obleser, J. (2015). Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. *Cortex*, *68*, 100–110.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Ladefoged, P., Silverstein, R., & Papcun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America*, *54*, 1105–1108.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice response. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 205–222). Hillsdale, NJ: Erlbaum.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Loucks, T., Kraft, S. J., Choo, A. L., Sharma, H., & Ambrose, N. G. (2011). Functional brain activation differences in stuttering identified with a rapid fMRI sequence. *Journal of Fluency Disorders*, *36*, 302–307.

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., et al. (2010). The neural substrates for atypical planning and execution of word production in stuttering. *Experimental Neurology*, *221*, 146–156.

Macey, P. M., Macey, K. E., Kumar, R., & Harper, R. M. (2004). A method for removal of global effects from fMRI time series. *Neuroimage*, *22*, 360–366.

McGuire, P. K., David, A. S., Murray, R. M., Frackowiak, R. S. J., Frith, C. D., Wright, I., et al. (1995). Abnormal monitoring of inner speech: A physiological basis for auditory hallucinations. *Lancet*, *346*, 596–600.

McLeod, P., & Posner, M. I. (1984). Privileged loops from percept to act. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 55–66). Hillsdale, NJ: Erlbaum.

Mehta, S., Grabowski, T. J., Razavi, M., Eaton, B., & Bolinger, L. (2006). Analysis of speech-related variance in rapid event-related fMRI using a time-aware acquisition system. *Neuroimage*, *29*, 1278–1293.

Nooteboom, S., & Quené, H. (2019). Temporal aspects of self-monitoring for speech errors. *Journal of Memory and Language*, *105*, 43–59.

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*, 403–410.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283–2289.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*, 528–537.
Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013). Attention for speaking: Domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience*, *7*, 832.
Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*, 97–132.
Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research*, *36*, 472–487.
Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, *21*, 150–164.
Severens, E., Kühn, S., Hartsuiker, R. J., & Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Social Cognitive and Affective Neuroscience*, *7*, 431–435.
Shapiro, A. K., Shapiro, E. S., Young, J. G., & Feinberg, T. E. (1988). *Gilles de la Tourette syndrome* (2nd ed.). New York: Raven Press.
Slevc, L. R., & Ferreira, V. S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language*, *54*, 515–540.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
Swick, D., & Chatham, C. H. (2014). Ten years of inhibition revisited. *Frontiers in Human Neuroscience*, *8*, 329.
van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584–589.
van den Wildenberg, W. P. M., & Christoffels, I. K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Frontiers in Psychology*, *1*, 145.
Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop signal and stop-change paradigms. *Neuroscience & Biobehavioral Reviews*, *33*, 647–661.
Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.
Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, *18*, 1923–1932.
Zaitsev, M., Hennig, J., & Speck, O. (2004). Point spread function mapping with parallel imaging techniques and high acceleration factors: Fast, robust, and flexible method for echo-planar imaging distortion correction. *Magnetic Resonance in Medicine*, *52*, 1156–1166.

Downloaded from http://direct.mit.edu/jocn/article-pdf/31/12/1946/1860995/jocn_a_01462.pdf by guest on 19 September 2021