

Research Article

Auditory Feedback Control Mechanisms Do Not Contribute to Cortical Hyperactivity Within the Voice Production Network in Adductor Spasmodic Dysphonia

Ayoub Daliri,^{a,b} Elizabeth S. Heller Murray,^a Anne J. Blood,^c James Burns,^c J. Pieter Noordzij,^d Alfonso Nieto-Castanon,^a Jason A. Tourville,^a and Frank H. Guenther^a

Purpose: Adductor spasmodic dysphonia (ADSD), the most common form of spasmodic dysphonia, is a debilitating voice disorder characterized by hyperactivity and muscle spasms in the vocal folds during speech. Prior neuroimaging studies have noted excessive brain activity during speech in participants with ADSD compared to controls. Speech involves an auditory feedback control mechanism that generates motor commands aimed at eliminating disparities between desired and actual auditory signals. Thus, excessive neural activity in ADSD during speech may reflect, at least in part, increased engagement of the auditory feedback control mechanism as it attempts to correct vocal production errors detected through audition.

Method: To test this possibility, functional magnetic resonance imaging was used to identify differences between participants with ADSD ($n = 12$) and age-matched controls ($n = 12$) in (a) brain activity when producing speech under different auditory feedback conditions and (b) resting-state functional connectivity within the cortical network responsible for vocalization.

Results: As seen in prior studies, the ADSD group had significantly higher activity than the control group during speech with normal auditory feedback (compared to a silent baseline task) in three left-hemisphere cortical regions: ventral Rolandic (sensorimotor) cortex, anterior planum temporale, and posterior superior temporal gyrus/planum temporale. Importantly, this same pattern of hyperactivity was also found when auditory feedback control of speech was eliminated through masking noise. Furthermore, the ADSD group had significantly higher resting-state functional connectivity between sensorimotor and auditory cortical regions within the left hemisphere as well as between the left and right hemispheres.

Conclusions: Together, our results indicate that hyperactivation in the cortical speech network of individuals with ADSD does not result from hyperactive auditory feedback control mechanisms and rather is likely related to impairments in somatosensory feedback control and/or feedforward control mechanisms.

Spasmodic dysphonia is a neurological voice disorder that is characterized by involuntary spasms of laryngeal muscles during speech production. The most common form of spasmodic dysphonia, impacting 80%–90% of people with the disorder, is the adductor type (adductor

spasmodic dysphonia [ADSD]) that is associated with spasms in closing muscles of the vocal folds, resulting in voice breaks (e.g., Ludlow, 2011; Schweinfurth et al., 2002). Generally, individuals with ADSD have a strained voice quality, and their voice is disrupted by intermittent voice breaks. ADSD is a task-specific disorder with symptoms that manifest primarily during vowels at the beginning or middle of words (e.g., *Anyway, I'll eat*; Erickson, 2003; Ludlow, 2011; Roy et al., 2005; Schweinfurth et al., 2002).

Although the pathophysiology of ADSD is not known, past studies have provided compelling evidence that the disorder is associated with abnormalities in sensorimotor processes (Battistella et al., 2016, 2018; Simonyan & Ludlow, 2010; Simonyan et al., 2008). Both structural and functional neuroimaging studies have reported that

^aDepartment of Speech, Language, & Hearing Sciences, Boston University, MA

^bCollege of Health Solutions, Arizona State University, Tempe

^cDepartment of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston

^dOtolaryngology, Head & Neck Surgery, Boston University School of Medicine, MA

Correspondence to Ayoub Daliri: Ayoub.daliri@asu.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Kate Bunton

Received February 13, 2019

Accepted November 5, 2019

https://doi.org/10.1044/2019_JSLHR-19-00325

Disclosure: The authors have declared that no competing interests existed at the time of publication.

ADSD is associated with abnormalities in key brain regions (cortical and subcortical) responsible for speech movement production (Ali et al., 2006; Haslinger et al., 2005; Simonyan & Ludlow, 2010). For example, functional neuroimaging studies of speech production in people with ADSA have reported hyperactivation in laryngeal and orofacial sensorimotor cortex including ventral premotor and motor regions as well as auditory and somatosensory cortices (Simonyan & Ludlow, 2010). It should be noted that, in studies of spasmodic dysphonia, the abnormalities in sensorimotor regions are more commonly evident as hyperactivation (Ali et al., 2006; Bianchi et al., 2017; Kirke et al., 2017; Kiyuna et al., 2017; Simonyan & Ludlow, 2010, 2012), although reduced activation in these regions has been reported (e.g., Haslinger et al., 2005). These neural characteristics of spasmodic dysphonia are broadly consistent with findings of studies of other types of focal dystonias. For example, studies of focal dystonia have reported abnormalities in the primary sensorimotor cortices and also higher order motor and associative cortical regions that resemble those of the spasmodic dysphonia (Lehéricy et al., 2013; Zoons et al., 2011). Furthermore, studies have reported abnormal functional connectivity between sensorimotor regions in individuals with ADSA (Battistella et al., 2016, 2018; Simonyan et al., 2013). Together, these studies suggest that ADSA is more commonly associated with overactivation in auditory, somatosensory, and motor regions that may interfere with normal sensorimotor processes. However, it is unclear whether the overactivation in these regions is related to malfunctioning auditory feedback, somatosensory feedback, or feedforward control mechanisms.

In this study, we specifically examined the contribution of auditory feedback control mechanisms to cortical hyperactivity in individuals with ADSA compared to control participants. Current models of speech production (Guenther, 2016; Hickok, 2012; Houde & Nagarajan, 2011) posit that the brain uses a combination of auditory feedback, somatosensory feedback, and feedforward mechanisms to accurately control the speech apparatus. For example, according to the Directions Into Velocities of Articulators model, the desired auditory output is compared to incoming auditory feedback during speech, and mismatches between the desired and actual auditory signals (e.g., formant frequencies, pitch, voice quality) will result in “error signals” in auditory cortex (Guenther, 2006, 2016; Tourville & Guenther, 2011). These signals in turn may cause the auditory feedback control subsystem, which operates in parallel with feedforward and somatosensory feedback control mechanisms, to generate motor commands that attempt to correct the perceived errors. It is, thus, possible that increased brain activity in the speech production network seen during speech in individuals with ADSA is due, at least in part, to this auditory feedback control mechanism. We tested this possibility by measuring brain activity in individuals with ADSA and matched controls while speaking under normal and noise-masked auditory feedback conditions. Specifically, we first used a localizer task to find brain regions involved in voice production and then examined between-group differences in

brain activity in the defined brain regions in two conditions: normal speaking and noise-masked speaking. If auditory feedback control mechanisms are responsible for excessive activity found during speech in participants with ADSA, then this excess activity compared to neurotypical controls should be reduced or eliminated when auditory feedback is masked with noise, given that noise masking will minimize or eliminate the detection of impaired vocalization through auditory feedback. In other words, we hypothesized that, if hyperactivity in the individuals with ADSA compared to neurotypical controls during normal speaking is due to the detection and attempted correction of auditory errors, then this between-group difference should be diminished when auditory errors can no longer be detected in the noise-masked condition. Alternatively, if a similar amount of excessive activity is found in participants with ADSA compared to controls for normal and noise-masked auditory feedback conditions, auditory feedback control mechanisms can be eliminated as a major source of excessive brain activity in ADSA. In other words, a similar between-group difference in normal speaking and noise-masked speaking conditions would suggest that the hyperactivation in the voice production network of individuals with ADSA does not result from hyperactive auditory feedback control mechanisms and rather is likely related to impairments in somatosensory feedback control and/or feedforward control mechanisms. In addition to investigating brain activity during speech, given past reports of abnormal functional connectivity in ADSA (Battistella et al., 2016, 2018; Simonyan et al., 2013), we also measured resting-state functional connectivity within the network of brain regions responsible for voicing in order to further characterize anomalies within the voice production network in ADSA. Specifically, we hypothesized that brain regions with increased activity during speaking in individuals with ADSA may also have abnormal functional connectivity with other regions in the voice production network.

Method

Participants

Participants were 12 patients with ADSA (seven women; $M_{\text{age}} = 54.17$ years, $SD_{\text{age}} = 9.91$) and 12 healthy volunteers (seven women; $M_{\text{age}} = 54.42$ years, $SD_{\text{age}} = 9.17$). All participants (a) were native speakers of American English; (b) were right-handed according to the 10-item version of the Edinburgh Handedness Inventory (laterality score larger than +40; Oldfield, 1971); (c) had no history of neurological, psychological, or communication disorders (other than ADSA in the patient group); and (d) had normal binaural hearing (pure-tone threshold ≤ 25 dB HL at octave frequencies of 500–4000 Hz). Participants with ADSA were diagnosed by an experienced laryngologist. All patients were fully symptomatic, and at least 3 months had passed since their last botulinum toxin treatments (note that information regarding head and neck dystonia, comorbidity, and

medication was not available for all patients). The average duration of ASD since the diagnosis was 11.25 years (range: 1–32 years). All participants were naïve to the purpose of the study and provided written informed consent prior to participation in the study. The institutional review board of Boston University approved the procedures of the study.

Procedure

The study consisted of two sessions. In the first session, participants were interviewed to ensure their qualifications according to the abovementioned inclusion criteria. In this session, the experimental tasks (see below for the description of the tasks) were explained to participants, and they were instructed to complete four runs of the experiment (see below for details) to become familiar with the experimental tasks. In the second session, participants completed two behavioral experiments while lying down inside a magnetic resonance imaging (MRI) scanner. In both experiments, visual stimuli and instructions were projected onto a screen viewed from within the scanner via a mirror attached to the head coil. Speech signals were transduced and amplified by a fiberoptic microphone (Fibersound Model FOM1-MR-30m) that was attached to the head coil; the signals were then sent to a Lenovo ThinkPad X61s, recorded (sampling rate = 16000 Hz), and replayed to the subject with a negligible delay (~18 ms). Auditory feedback was amplified (Behringer 802) and delivered through MRI-compatible insert earphones (Sensimetrics Model S-14). The feedback gain was calibrated prior to each experiment such that 70 dB SPL at the microphone resulted in 75 dB SPL at the earphones. In addition to a structural scan for image registration and activity localization, participants performed two tasks in the scanner: a localizer task and a sentence production task.

Functional MRI Localizer Task

A simple voicing task was used to identify regions involved in voice production using a continuous sampling technique. The localizer task was not intended to elicit symptoms in individuals with ASD (participants with ASD were asymptomatic during this task). Participants completed two runs of the localizer task. In each run, participants completed 10 trials of a baseline condition (paced breathing) and 10 trials of a voicing condition (paced voicing). The order of trials was randomized within each run. Each trial was preceded by a silent interval (1.5–2.5 s) during which participants were visually instructed about the upcoming task to be performed. Each trial lasted approximately 12 s. In the baseline condition, participants were instructed to inhale through their nose, keeping their mouth closed, once every other second while watching a crosshair on the screen. In the voicing condition, participants were instructed to “hum” in a monotone voice once every second while watching a crosshair on the screen. A

contrast of “voicing–baseline” was used to identify voicing-related regions in each participant.

Sentence Production Task

In this task, we examined activation of regions involved in speech production with and without noise-masked auditory feedback. Participants completed four runs of the sentence production task. During this experiment, in combination with behavioral speech data, we collected functional imaging data using an event-triggered sparse sampling technique. This allowed participants to receive auditory feedback related to their production in the absence of scanner noise (Hall et al., 1999). The experiment consisted of three conditions¹: “baseline,” “normal speaking,” and “speaking under masking noise.” A sparse sampling scanning protocol (described below) was utilized to eliminate scanner noise while the participants were speaking. In each run, participants completed 12 randomly distributed trials of each condition (36 trials in each run). Each trial started with a presentation of a visual stimulus (sentence or non-linguistic symbols). The stimulus stayed on the screen for 3.5 s and disappeared afterward. Each trial lasted 8 s. In the baseline condition, visual stimuli were nonlinguistic symbols, and participants were instructed (prior to scanning) to watch the screen without producing any movements or sounds. In the normal speaking condition, visual stimuli were sentences (“Anyway, I’ll eat”; “Anyway, I’ll argue”; “Anyway, I’ll iron”). The sentences were purposefully designed to elicit the symptoms of the adductor type of spasmodic dysphonia (i.e., muscle spasms during vowels) in the ASD group (Ludlow, 2011). Participants overtly produced the sentences while they received auditory feedback related to their production (through insert earphones). Participants were instructed to read the sentences as soon as they appeared on the screen and as consistently as possible so that they maintained their natural intonation pattern, rhythm, and loudness throughout the study. In the “speaking under masking noise” condition, participants overtly produced the sentences while their auditory feedback was masked by speech-modulated noise (5 dB SPL greater than the participant’s speech output; see Procedure section). It should be noted this procedure is different from the noise-masking procedure used in previous studies (Christoffels et al., 2011; Kleber et al., 2017). In previous studies, the masking noise was applied throughout a production trial with a constant intensity, whereas (a) we applied the masking noise when the subject was producing speech and no noise was applied when the participant was not producing speech, and

¹In addition to these three conditions, the experiment included another condition in which fundamental frequency of speech was altered in real time. Although this condition was modeled in the first-level analysis of the single subject and the general linear models, it was dropped from all group analyses, as no behavioral response to the feedback perturbation was evident. It should be noted that the pattern of group differences during this removed condition was very similar to those in the normal feedback and noise-masked conditions.

(b) the noise amplitude was modulated by the speech envelope.

MRI Image Acquisition

MRI images were acquired using a Siemens Magnetom Skyra 3T scanner equipped with a 32-channel phased array head coil in a single imaging session at MGH/HST Martinos Center for Biomedical Imaging. We collected four different types of data: (a) high-resolution T1-weighted anatomical scan, (b) sparse-sampled T2*-weighted functional scans, (c) continuous-sampled T2*-weighted functional scans, and (d) continuous-sampled T2*-weighted resting-state functional scans.

Prior to functional runs, a whole-head, high-resolution T1-weighted image was collected (magnetization prepared rapid gradient echo sequence; repetition time (TR) = 2530 ms, echo time (TE) = 1.69 ms, inversion time (TI) = 1100 ms, flip angle = 7°, voxel resolution = 1.0 mm × 1.0 mm × 1.0 mm, field of view (FOV) = 256 × 256, 176 sagittal slices). During the localizer task, two functional runs were collected using continuous-sampling technique (echo-planar imaging [EPI]; 101 measurements; TR = 2.8 s, TE = 30 ms, flip angle = 90°, voxel resolution = 3.0 mm × 3.0 mm × 3.0 mm, FOV = 64 × 64, 46 transverse slices).

During the sentence production task, four functional runs were collected using even-related sparse-sampled T2*-weighted gradient EPI scans (TR = 8.0 s, acquisition time (TA) = 2.5 s, delay = 5.5 s, TE = 30 ms, flip angle = 90°, voxel resolution = 3.0 mm × 3.0 mm × 3.0 mm, FOV = 64 × 64, 46 transverse slices). The functional images were automatically registered to the anterior commissure–posterior commissure (AC-PC) line, and the acquisition of the scans was automatically triggered by the onset of the visual stimuli. To ensure the stabilization of longitudinal magnetization, one additional volume was collected prior to each functional run.

Additionally, we collected one run of resting-state functional data (~6 min). During this run, participants were instructed to keep their eyes closed, relax, and move as little as possible. We collected 315 functional images using simultaneous multislice EPI (315 measurements; TR = 1.13 s, TE = 30 ms, flip angle = 60°, voxel resolution = 3.0 mm × 3.0 mm × 3.0 mm, FOV = 72 × 72, 51 transverse slices, simultaneous multislice factor = 3).

Data Analyses

Acoustic Data Analysis and Severity Measure

Acoustic analyses were carried out using custom-written scripts for Praat (Boersma, 2002) and MATLAB. We used sentence productions of participants in the normal speaking condition from the first session—four runs, with 12 sentences in each run. Under noise-masked speaking conditions, speakers typically increase their speech intensities. To ensure that the two groups have similar patterns of increase in intensity, we measured overall intensity of each participant and compared the two groups using a

repeated-measures analysis of variance with condition (normal speaking vs. noise-masked speaking) as the within-subject variable and group as the between-group variable. As expected, the main effect of condition was statistically significant, $F(1, 22) = 23.614, p < .001$, with higher intensity in the noise-masked condition in comparison with the normal speaking condition. We did not find a statistically significant effect of group, $F(1, 22) = 0.14, p = .710$, or a Group × Condition interaction, $F(1, 22) = 0.586, p = .452$, suggesting that the noise-masking condition similarly influenced both groups.

A certified speech-language pathologist performed subjective auditory–perceptual ratings on speech samples from Session 1, rating the vocal attribute of overall severity with an electronic version of the clinically available tool, the Consensus Auditory-Perceptual Evaluation of Voice (Kempster et al., 2009). For each participant, the perceptual rating was determined after listening to the participant's first few speech samples. These chosen samples were from the initial portion of the experimental session in order to avoid any vocal changes that could have occurred due to multiple repetitions of the target sentences. Auditory–perceptual ratings were made by moving a slider along a 100-mm horizontal line, with nonlinearly spaced anchors of “mildly deviant,” “moderately deviant,” and “severely deviant” written beneath the horizontal line. The position of the slider from the leftmost portion of the line was measured in millimeters. The scale ranged from 0 = *normal voice* to 100 = *extremely deviant voice* (Kempster et al., 2009). The severity score ranged from 14 to 91 ($M = 43.7, SD = 26.9$) for the ADSD group and from 0 to 7.9 ($M = 1.4, SD = 2.1$) for the control group. It should be noted that we did not perform interrater reliability, and this may have influenced the robustness of the severity scores.

Task-Based Functional Imaging Analyses

Preprocessing

Preprocessing of the imaging data was conducted using the Nipype neuroimaging software interface (Version 0.14.0; Gorgolewski et al., 2011), which provides a Python-based interface to create pipelines by combining and using algorithms from different neuroimaging software packages. The analyses were conducted on a Linux-based, high performance–computing cluster.

The high-resolution T1-weighted images were entered into the FreeSurfer software package (Version 5.3; <http://surfer.nmr.mgh.harvard.edu>; Dale et al., 1999) with default parameters to perform skull stripping, image registration, image segmentation, and cortical surface reconstruction. The reconstructed surfaces for each individual were visually inspected to ensure the accuracy and quality of the reconstruction. Functional volumes were motion corrected and coregistered to the individual's anatomical volume (preprocessed in FreeSurfer). Because functional volumes of the localizer task were continuously sampled (as opposed to sparse sampled), slice-time corrections were applied to the functional time series of the localizer task. Additionally,

functional volumes with (a) intensity higher than 3 *SDs* above the individual's mean intensity (calculated for each run separately) or (b) motions greater than 1 mm were marked as outliers using Artifact Detection Tools (http://www.nitrc.org/projects/artifact_detect). Outliers were entered into the first-level design matrix as nuisance regressors (one column per outlier). The motion-corrected and co-registered functional time series were temporally high-pass filtered (128-s cutoff frequency) and spatially smoothed on the cortical surface with a Gaussian filter (6-mm full width at half maximum).

For each participant and each condition of interest (in both sentence production and localizer tasks), a train of impulses—representing stimulus events—was convolved with a canonical hemodynamic response function to generate a simulated blood oxygen level-dependent (BOLD) response. The first-level model was estimated using a general linear model with (a) task regressors—simulated BOLD responses in all three main conditions (baseline, normal speaking, and speaking under masking noise) and the condition that was dropped from the group analysis (i.e., fundamental frequency manipulation), (b) motion regressors (three translation and three rotation parameters), and (c) nuisance regressors (outlier volumes identified in preprocessing). For the localizer task, we calculated one contrast based on the voicing condition (i.e., humming) and the baseline condition (i.e., breathing) in this task. For the sentence production, we calculated three contrasts based on the normal speaking, speaking under masking noise, and no-speaking baseline conditions in this task (i.e., normal speaking vs. baseline, speaking under masking noise vs. baseline, and speaking under masking noise vs. normal speaking). Note that each task (localizer and sentence production) had its own baseline condition.

Functional Region-of-Interest Definition

To address the negative impact of interindividual anatomical variability on sensitivity of results in group analyses (Nieto-Castañón & Fedorenko, 2012), we conducted two group-constrained subject-specific analyses (Fedorenko et al., 2010; Julian et al., 2012). The goal of this analysis was to determine a set of voice-sensitive regions of interest (ROIs) using the preprocessed functional data in the localizer task. For this purpose, cortical surfaces were extracted from the normalized (to Montreal Neurological Institute template) individual activation maps for the “voicing–baseline” contrasts of all subjects across the two groups, liberally thresholded (uncorrected $p < .05$); vertices with $p > .05$ were assigned the value of 0 (i.e., inactive), and those with $p < .05$ were assigned the value of 1 (i.e., active). Note that activation maps are the change in the BOLD signal in the voicing condition relative to the baseline condition in the localizer task. The new thresholded activation map for all individuals were overlaid and averaged in each voxel to create a probabilistic overlap map (each voxel had a value in the range of 0–1). Each voxel of the overlap map contained the percentage of subjects with suprathreshold voicing-related activation in that voxel. Then, a watershed algorithm (implemented in MATLAB;

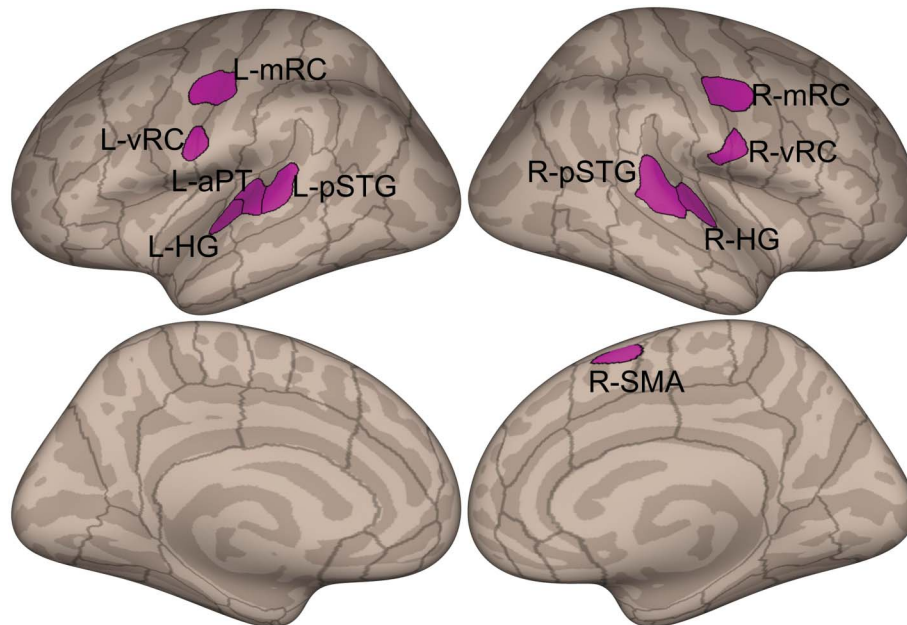
for detailed descriptions of the watershed method, see Nieto-Castañón & Fedorenko, 2012) was used to parcellate the probabilistic overlap map into distinct ROIs. Cortical ROIs that consisted of at least 300 vertices for at least 60% of all subjects (more than 14 subjects out of all 24 subjects) were considered the final cortical ROIs (Nieto-Castañón & Fedorenko, 2012). This analysis resulted in 10 cortical ROIs (five in each hemisphere; see Figure 1). The second analysis was a volume-based analysis, focusing on subcortical activation. However, this analysis did not result in significant subcortical ROIs and will not be discussed further. It should be noted that the voicing–baseline contrast (i.e., humming vs. breathing) detects ROIs that are involved in voice production and processing and may not detect all regions that are involved in speech production and processing.

ROI-to-ROI Resting-State Functional Imaging Analyses

We used the CONN Functional Connectivity Toolbox (v.17.f; <https://www.nitrc.org/projects/conn>; Whitfield-Gabrieli & Nieto-Castañón, 2012) to perform spatial and temporal preprocessing and to compute a functional correlation matrix for each subject. Spatial preprocessing steps included slice-timing correction (to correct for acquisition time), realignment (to correct for interscan head motions), structural–functional coregistration (using subject's structural T1-weighted image and the mean functional image), segmentation of functional and structural images (gray matter, white matter, and cerebrospinal fluid), spatial smoothing (6-mm full width at half maximum), and normalization to the Montreal Neurological Institute coordinate space. The CONN toolbox uses SPM12 (<https://www.fil.ion.ucl.ac.uk/spm>) to perform all spatial preprocessing steps. In addition, the Artifact Detection Toolbox (http://www.nitrc.org/projects/artifact_detect/) was used to identify outlier functional volumes (displacement > 1 mm and intensity z threshold > 3 *SDs* from the mean). To reduce the impact of motion and physiological noise factors and to improve the overall validity and robustness of results, several temporal preprocessing steps were used. We used covariate regression analysis to remove the effects of (a) 12 movement parameters (six rigid-body movement parameters and their first derivatives), (b) principal components of subject-specific white matter and cerebrospinal fluid, and (c) outlier volumes. Finally, we temporally filtered the time series (0.008–0.15 Hz) and linearly detrended the time series.

We examined the resting-state data using ROI-to-ROI functional connectivity analysis. For this purpose, correlation matrices were computed by calculating Pearson correlation coefficients between time series of pairs of the functional ROIs calculated using the group-constrained subject-specific analyses method (see Functional Region-of-Interest Definition section). Prior to entering to group-level statistical analyses, Fisher's transformation was applied to the bivariate correlation coefficients to improve the normality assumptions (see below for statistical analyses of ROI-to-ROI connectivity measures).

Figure 1. Functional regions of interest derived from the voicing (humming) vs. baseline (breathing) contrast of the functional localizer task plotted on inflated cortical surfaces. Upper left: left hemisphere lateral view. Upper right: right hemisphere lateral view. Lower left: left hemisphere medial view. Lower right: right hemisphere medial view. Dark shaded areas indicate sulci. Lines indicate boundaries between anatomical regions. The approximate anatomical locations of the left hemisphere regions of interest are as follows: L-mRC = mid-Rolandic cortex; L-HG = lateral Heschl's gyrus; L-vRC = ventral Rolandic cortex; L-aPT = anterior planum temporale; L-pSTG = posterior superior temporal gyrus/planum temporale. The approximate anatomical locations of the right hemisphere regions of interest are as follows: R-mRC = mid-Rolandic cortex; R-HG = Heschl's gyrus; R-pSTG = posterior superior temporal gyrus/planum temporale; R-vRC = ventral Rolandic cortex; R-SMA = supplementary motor area/pre-SMA.



Statistical Analysis

The cortical functional ROIs derived from the localizer task (i.e., brain regions involved in voice production) were used (a) to extract activation maps in each of the contrasts in the sentence production task and (b) as seed ROIs to study ROI-to-ROI functional connectivity based on the resting-state data.

As a dependent measure for the sentence production task, we calculated change in the BOLD signal for a given contrast (normal speaking vs. baseline, noise-masked speaking vs. baseline, and noise-masked speaking vs. normal speaking) in all vertices within each of the 10 ROIs. Then, we averaged the values across all vertices of each ROI to estimate the overall change in the BOLD signal in each ROI of a given contrast. For each contrast, we used a linear mixed-effects model implemented in the lme4 package (Bates et al., 2015) with Group (ADSD and control) and ROI (10 levels) as fixed factors and subject as a random factor (random intercept). To determine statistical significance, we used the lmerTest package (Kuznetsova et al., 2017) with the Satterthwaite's method for estimating degrees of freedom. We used the emmeans package (Lenth, 2019) to conduct post hoc between-group comparisons with the false discovery rate (FDR) method of correction for multiple comparisons. We used the Kenward-Roger method to determine the degrees of freedom of the post hoc tests.

Given that we were interested in between-group differences, post hoc tests were limited to comparisons between groups in different ROIs (i.e., different ROIs were not compared together). These statistical analyses were conducted in R Version 3.5.1 (R Core Team, 2018).

As mentioned above, the determined ROIs based on the localizer task were used to perform ROI-to-ROI functional connectivity on the resting-state data. For this purpose, we examined the entire matrix of ROI-to-ROI connections across all 10 ROIs. To correct for multiple comparisons, we used the "FDR analysis-level correction" ($p < .05$) option available in the CONN toolbox (Whitfield-Gabrieli & Nieto-Castañón, 2012). In this method, FDR multiple comparison correction was applied across all individual functional connections included in the analysis (there were $N \times (N - 1) / 2$ connections in the analysis of bivariate connectivity measures among $N = 10$ ROIs). In other words, between-group differences were examined after applying FDR correction across the entire analysis (accounting for the number of ROIs included in the analysis as seed and target ROIs). To examine potential relationships between severity scores of the ADSD group and neural measures (brain activation in ROIs with significant between-group differences and significant functional connectivity), we calculated Pearson correlation coefficients using the R package Psych (Revelle, 2018).

Results

Functional Localizer

Figure 1 shows the ROIs derived from the functional localizer task (“voicing–baseline” contrast) using the group-constrained watershed segmentation. A total of 10 ROIs, five in each cortical hemisphere, were identified. The approximate anatomical locations of the left hemisphere ROIs are (1) mid-Rolandic cortex (mRC), (2) lateral Heschl’s gyrus, (3) ventral Rolandic cortex (vRC), (4) anterior planum temporale (aPT), and (5) posterior superior temporal gyrus/planum temporale (pSTG). Regions 1 and 3 are located within the primary sensorimotor cortex (spanning primary motor and somatosensory regions in the precentral and postcentral gyrus, respectively), whereas Regions 2, 4, and 5 are in the auditory cortex. The approximate anatomical locations of the right hemisphere ROIs are (6) mRC, (7) Heschl’s gyrus, (8) pSTG, (9) vRC, and (10) supplementary motor area/presupplementary motor area.

Sentence Production

Examining the “normal speech versus baseline” contrast, we found a statistically significant main effect of ROI, $F(9, 216) = 8.903, p < .001$; a significant Group \times ROI interaction, $F(9, 216) = 3.905, p < .001$; and a nonsignificant main effect of Group, $F(1, 24) = 3.532, p = .072$. Similarly, examining the “noise-masked speech versus baseline” contrast, we found a statistically significant main effect of ROI, $F(9, 216) = 10.595, p < .001$; a significant Group \times ROI interaction, $F(9, 216) = 3.928, p < .001$; and a nonsignificant main effect of Group, $F(1, 24) = 3.289, p = .083$. Our analysis of the “noise–masked speech versus normal speech” contrast resulted in a statistically significant main effect of ROI, $F(9, 216) = 5.165, p < .001$; a nonsignificant Group \times ROI interaction, $F(9, 216) = 1.119, p = .351$; and a nonsignificant main effect of Group, $F(1, 24) = 0.632, p = .434$.

Given the hypotheses under investigation, we conducted a series of post hoc tests with the focus on between-group comparisons in different ROIs for contrasts with statistically significant Group \times ROI interaction. Note that whether activation in one ROI was different from that in another ROI was not the focus of this study, and therefore, we did not perform post hoc tests to examine between-ROI differences. As illustrated in Figure 2, for both the “normal speech versus baseline” contrast and “noise-masked speech versus baseline” contrast of the sentence production functional MRI task (with a statistically significant Group \times ROI interaction), participants with ASD showed significantly greater activity than control participants ($p < .05$, FDR-corrected across 10 ROIs) in three left hemisphere regions: L-vRC, L-aPT, and L-pSTG. No right hemisphere ROIs showed significant activity differences in any of the speech conditions contrasted with baseline. Overall, very similar activation patterns were observed for both contrasts.

In our main analysis, all significant between-group differences were in ROIs located in the left hemisphere. Therefore, we conducted a post hoc analysis to examine the group \times hemisphere interaction. For this purpose, for each contrast, data in all ROIs of each hemisphere were collapsed and entered in a linear mixed-effects model with group (ASD and control) and hemisphere (right and left) as fixed factors and subject as a random factor (random intercept). Examining the “normal speech versus baseline” contrast, we found a statistically significant main effect of hemisphere, $F(1, 216) = 9.766, p = .002$, and a significant Group \times Hemisphere interaction, $F(1, 216) = 14.660, p < .001$. Similarly, examining the “noise-masked speech versus baseline” contrast, we found a statistically significant main effect of hemisphere, $F(1, 216) = 6.471, p = .011$, and a significant Group \times Hemisphere interaction, $F(1, 216) = 9.346, p = .003$. As shown in Figures 3A and 3B, these Group \times Hemisphere interactions indicated that the ASD group had larger activity in the left hemisphere relative to the right hemisphere ($p < .05$ FDR corrected) and the left hemisphere activity of the control group ($p < .05$ FDR corrected). Our analysis of the “noise-masked speech versus normal speech” contrast did not result in significant Group effect ($p = .434$), Hemisphere effect ($p = .325$), or Group \times Hemisphere interaction ($p = .190$).

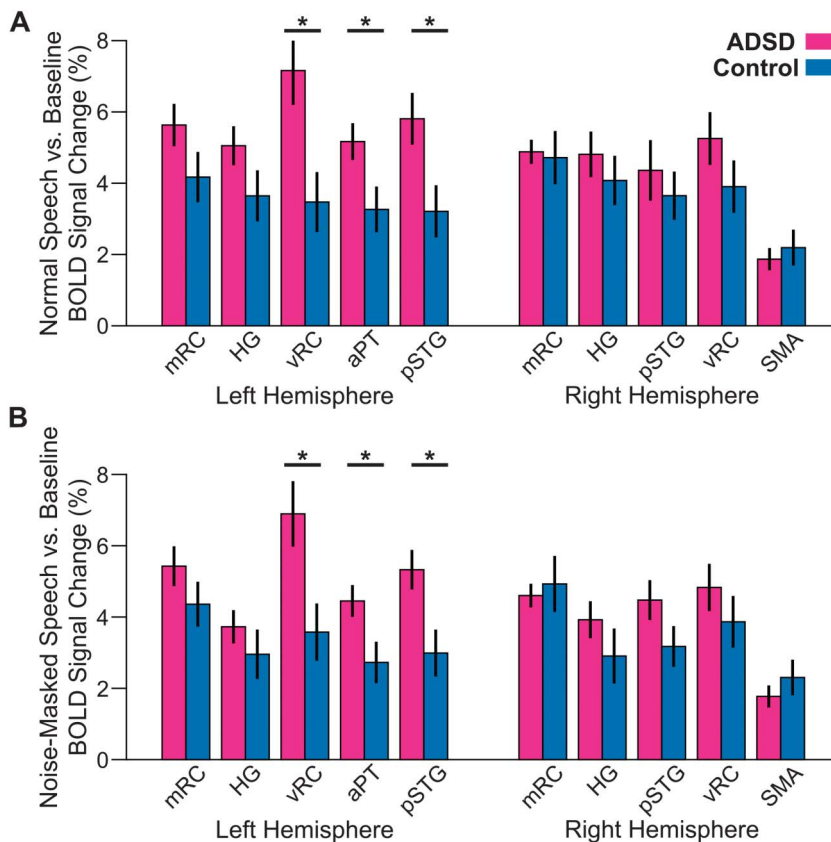
To examine potential relationships between severity scores of the ASD group and their brain activity, we conducted Pearson correlation coefficients for two contrasts that showed between-group differences. For the “normal speech versus baseline” contrast, we found that the severity scores did not correlate with brain activation in any of the three ROIs in which the ASD group had hyperactivation relative to the control group (L-vRC: $r = .310, p = .327$; L-aPT: $r = .508, p = .092$; L-pSTG: $r = -.144, p = .654$). Similarly, the severity scores did not correlate with any of the three ROIs of the “noise-masked speech versus baseline” contrast (L-vRC: $r = .112, p = .712$; L-aPT: $r = .110, p = .734$; L-pSTG: $r = .465, p = .128$).

Functional Connectivity

We used the ROIs defined based on the localizer task to conduct ROI-to-ROI connectivity analysis. Blue lines in Figure 4 indicate connections with significantly higher ($p < .05$, FDR analysis-level FDR-corrected for the number of ROIs included in the analysis as seed and target ROIs) functional connectivity for the ASD group compared to the control group. ROI-to-ROI resting-state functional connectivity was significantly stronger ($p < .05$, FDR analysis-level FDR-corrected) for the ASD group compared to the control group between the following ROI pairs: L-mRC and left Heschl’s gyrus, L-mRC and L-pSTG, L-vRC and L-pSTG, L-mRC and right Heschl’s gyrus, and R-mRC and L-pSTG.

To examine whether these higher ROI-to-ROI connections (five significant ROI-to-ROI connections) in the ASD group are correlated with the single-rater subjective

Figure 2. Region of interest activity for the normal speaking (A) and noise-masked speaking (B) conditions contrasted with the baseline condition in the sentence production functional magnetic resonance imaging task for participants with ASD (magenta) and control participants (blue). Significant group differences ($p < .05$, false discovery rate–corrected) are indicated by asterisks. See Figure 1 for region of interest definitions. Error bars correspond to standard error. BOLD = blood oxygen level dependent; ASD = adductor spasmodic dysphonia; mRC = mid-Rolandic cortex; HG = Heschl’s gyrus; vRC = ventral Rolandic cortex; aPT = anterior planum temporale; pSTG = posterior superior temporal gyrus; SMA = supplementary motor area.



measures of severity, we calculated Pearson correlation coefficients. Our analyses did not reveal any statistically significant relationship between the ROI-to-ROI connectivity and the severity score ($p > .093$ in all cases).

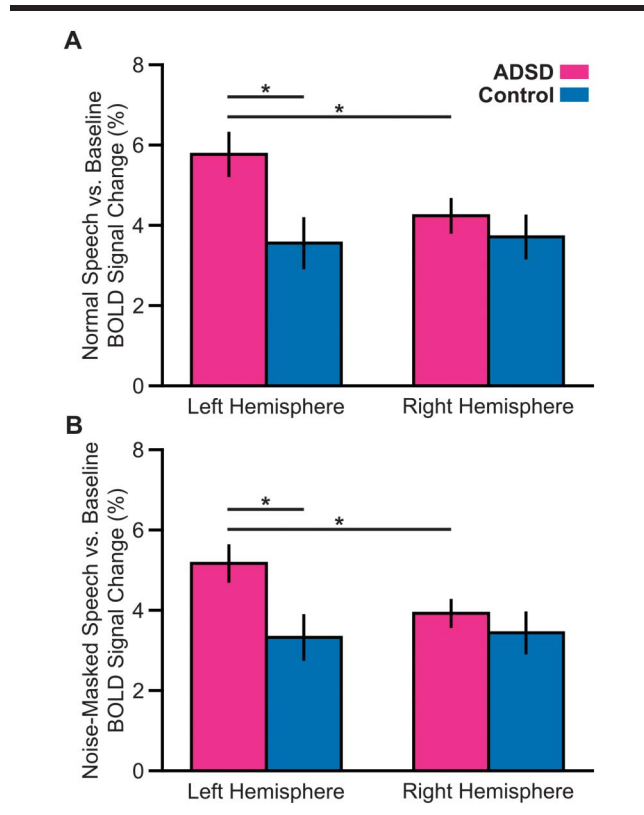
Discussion

Previous studies have reported hyperactivity in laryngeal and orofacial sensorimotor regions in people with ASD relative to healthy individuals (Ali et al., 2006; Hirano et al., 2001; Kiyuna et al., 2014, 2017; Simonyan & Ludlow, 2010, 2012; but see Haslinger et al., 2005). For example, Simonyan and Ludlow (2010) measured brain activation of individuals with ASD during a syllable production task and reported hyperactivity in bilateral ventral primary motor and somatosensory cortices, superior and medial temporal gyri, parietal operculum, and several subcortical regions. It should be noted that functional abnormalities in laryngeal sensorimotor and auditory regions have been observed during both symptomatic and nonsymptomatic speech tasks (Bianchi et al., 2017; Simonyan & Ludlow, 2012; Simonyan et al., 2008). Together, these results have

led researchers to suggest that abnormal feedforward control mechanisms and/or somatosensory feedback control mechanisms may underlie this excess activity and associated behavioral characteristics of ASD (Ali et al., 2006; Ludlow, 2011; Simonyan & Ludlow, 2010).

In the context of the Directions Into Velocities of Articulators model of speech production (e.g., Guenther, 2016), as well as other current speech production models that utilize a combination of auditory feedback control, somatosensory feedback control, and feedforward control mechanisms (Hickok, 2012; Houde & Nagarajan, 2011), the impaired voice quality evident in ASD speech may trigger auditory feedback control mechanisms in an attempt to correct the aberrant voice signal. The involvement of auditory feedback control mechanisms might thus contribute to the hyperactivity seen in neuroimaging studies of ASD as well as the excess muscle activation characteristic of the disorder. Here, we tested this hypothesis by imaging speakers with ASD and matched controls during the production of speech in different auditory feedback conditions that were designed to isolate contributions of auditory feedback control mechanisms.

Figure 3. Motivated by our main analysis (see Figure 2), we conducted a post hoc analysis to examine the Group \times Hemisphere interaction. Therefore, for each contrast, data in all regions of interest of each hemisphere were collapsed. Overall, for both contrasts (A: normal speech vs. baseline; B: noise-masked speech vs. baseline), we found that the ASD group had hyperactivation in the left hemisphere relative to the right hemisphere. We also found that the ASD group had hyperactivation in the left hemisphere relative to the control group (asterisk corresponds to $p < .05$ false discovery rate—corrected; see Figure 1 for region of interest definitions). Error bars correspond to standard error. BOLD = blood oxygen level dependent; ASD = adductor spasmodic dysphonia.



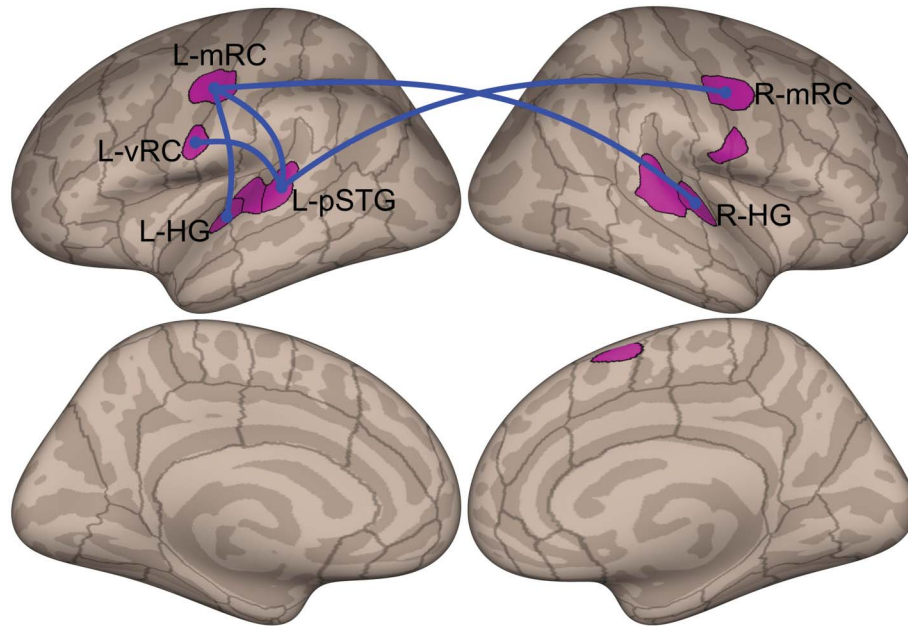
Consistent with previous studies of spasmodic dysphonia, we found that, under normal auditory feedback conditions, individuals with ASD had significantly higher activity than control subjects during symptomatic sentence production (compared to a silent baseline task) in three left hemisphere cortical regions: ventral Rolandic (sensorimotor) cortex, aPT, and pSTG. Interestingly, we did not find significant hyperactivation in a more dorsal sensorimotor cortical region that has been termed “laryngeal motor cortex” (LMC; Brown et al., 2008; Simonyan & Horwitz, 2011), corresponding to Regions 1 (L-mRC) and 6 (R-mRC) in the current study (see Figure 1). Like the current study, a number of prior studies have reported distinct laryngeal representations in ventral sensorimotor cortex (Guenther, 2016; Olthoff et al., 2008; Simonyan, 2014; Simonyan & Horwitz, 2011; Terumitsu et al., 2006). The finding of hyperactivation in the more ventral region in ASD, rather than the so-called LMC, speaks against the notion that the LMC is the sole locus of vocal control during speech,

suggesting instead that both the dorsal and ventral regions identified in the current study are heavily involved in vocalization. It should be noted that our results are also in large agreement with results of studies of other types of focal dystonia that have reported increased activity in sensorimotor cortices (Lehéricy et al., 2013; Zoons et al., 2011).

Critically, regarding the primary question of whether auditory feedback control mechanisms contribute substantially to left hemisphere cortical hyperactivity seen in ASD during speech, we found the same pattern of hyperactivity in participants with ASD compared to controls in the noise-masked speech condition. Since participants could not hear their vocal errors (if any) during noise-masked speech, we can conclude that left hemisphere cortical hyperactivity in participants with ASD is not caused by corrective motor commands from the auditory feedback control system in an attempt to correct vocal errors—the hyperactivity persists even when participants with ASD cannot hear their (error-prone) vocal output. By eliminating auditory feedback control as a major contributor to left hemisphere cortical hyperactivity in ASD, this finding lends further support to two alternative theoretical views, namely, that ASD arises from anomalies in the feedforward control system and/or anomalies in the somatosensory feedback control subsystem. It should be noted that we used a noise-masking procedure similar to our previous study (Ballard et al., 2018) in which the masking noise was applied when the subject was producing speech and the noise amplitude was modulated by the speech envelope. Although one might think of masking noise as a form of “auditory error,” it is important to note that the auditory errors dealt with by the auditory feedback control system for voice are errors in perceptually relevant acoustic parameters such as pitch or harmonic-to-noise ratio—the auditory feedback control system becomes less and less engaged as auditory feedback becomes less “natural sounding,” delayed, or less relevant (Daliri & Dittman, 2019; Daliri & Max, 2018; Liu & Larson, 2007; MacDonald et al., 2010; Max & Maffett, 2015; Mitsuya et al., 2017). For example, Liu and Larson (2007) showed that vocal responses to large pitch shifts are smaller than those to smaller pitch shifts, suggesting that the brain estimates the relevance of auditory errors and responds less when auditory errors are unnaturally large. Since key vocal parameters such as pitch and voice quality cannot be detected during noise masking, the auditory feedback control of these parameters was essentially disabled.

In addition, we found that the ASD group had significantly higher resting-state functional connectivity between sensorimotor and auditory cortical regions within the left hemisphere as well as between the left and right hemispheres. Prior studies have reported both abnormal functional and abnormal structural connectivity between sensorimotor regions in individuals with ASD (Battistella et al., 2016, 2018; Bianchi et al., 2017; Kiyuna et al., 2017; Simonyan et al., 2008). Most relevant to this study, previous studies that have examined abnormalities in functional connectivity of ASD have reported both increased and

Figure 4. ROI-to-ROI resting-state functional connectivity results. We used the ROIs defined based on the localizer task to conduct ROI-to-ROI connectivity analysis. Blue lines indicate connections with significantly higher functional connectivity for the adductor spasmodic dysphonia group compared to the control group ($p < .05$, false discovery rate analysis level corrected for the number of ROIs included in the analysis as seed and target ROIs). See Figure 1 for ROI definitions. ROI = region of interest; L = left; R = right; mRC = mid-Rolandic cortex; vRC = ventral Rolandic cortex; pSTG = posterior superior temporal gyrus; HG = Heschl's gyrus.



decreased functional connectivity in individuals with ASD relative to control participants (Battistella et al., 2016, 2017; Kiyuna et al., 2017). Our results are consistent with results of Kiyuna et al. who also used seed-based resting-state functional connectivity and reported that ASD is associated with increased functional connectivity in several cortical and subcortical regions. Most notably, they found increased functional connectivity between the motor cortex (right precentral gyrus) and the auditory-associated cortices (left middle and inferior temporal gyri), between the somatosensory cortex (left postcentral gyrus) and the frontal lobe (right frontal pole), and between the left inferior operculum and the right precentral and postcentral gyri. Overall, the increased functional connectivity between motor areas and auditory areas found in the current study (and previous studies) may arise due to chronic hyperactivation of these areas simultaneously during speech, resulting in Hebbian learning between the two areas. Alternatively, it may reflect an increased influence of efference copy activity from motor/premotor areas to auditory cortical areas as a result of hyperactivity in the former. Both of these explanations are consistent with either a somatosensory feedback control impairment or a feedforward control impairment as the cause of ASD.

One aspect of our results that deserves a comment is the lack of between-condition differences in any of the ROIs. For example, previous studies have reported increased brain activity (especially in auditory regions) in noise-masking conditions relative to normal auditory

feedback in neurotypical participants (Christoffels et al., 2011; Kleber et al., 2017). One potential explanation for this discrepancy between our results and those of previous studies is that we used specific ROIs that were related to voice production (based on the localizer task). For example, regions that would show between-condition differences may have been, at least partially, located outside the voice production ROIs used in this study (therefore, our analysis would not be able to detect them). It is possible that a whole-brain analysis (as opposed to ROI-based analysis) may have detected between-condition differences. Another potential explanation is related to the nature of the masking noise itself. In previous studies that reported increased brain activity when speech is masked by noise, the noise masking was applied throughout a production trial (Christoffels et al., 2011; Kleber et al., 2017). In contrast, our procedure only provided masking noise when the subject was actually vocalizing (the noise amplitude was modulated by the actual sound envelope produced by the speaker). Therefore, it is possible that the procedures (i.e., small but specific voice-related ROIs and speech-modulated noise masking) that we used in our study may have reduced the effects of noise masking. Given that the focus of our analyses was on the voicing-related ROIs, it is possible that there are potential group differences in regions outside the voicing-related ROIs. Finally, it should be noted that participants repeatedly read the sentences over the course of the sentence production task. It is possible that the repetitive nature of the task may have introduced

repetition effects that could potentially weaken brain responses.

In summary, we examined whether auditory feedback control mechanisms contribute to previously reported increased brain activity in the speech production network of individuals with ASD. We used functional MRI to identify differences between participants with ASD and age-matched controls in (a) brain activity when producing speech under different auditory feedback conditions and (b) resting-state functional connectivity within the cortical network responsible for vocalization. In the normal speaking condition, individuals with ASD had hyperactivation compared to controls in three left hemisphere cortical regions: ventral Rolandic (sensorimotor) cortex, aPT, and pSTG. Importantly, the same pattern of hyperactivity was evident in the noise-masked condition, in which on-line auditory feedback control is eliminated. Additionally, the ASD group had significantly higher resting-state functional connectivity between sensorimotor and auditory cortical regions within the left hemisphere as well as between the left and right hemispheres. Together, our results indicate that hyperactivity in the vocal production network of individuals with ASD does not result from hyperactive auditory feedback control mechanisms and rather is likely caused by impairments in somatosensory feedback control and/or feedforward control mechanisms.

Acknowledgments

This study was supported by National Institutes of Health Grants R01 DC002852 (F. Guenther, PI) and R21 DC017563 (A. Daliri, PI). We thank Elisa Golfopoulos for her contribution to subject recruitment and data collection of this project.

References

- Ali, S. O., Thomassen, M., Schulz, G. M., Hosey, L. A., Varga, M., Ludlow, C. L., & Braun, A. R. (2006). Alterations in CNS activity induced by botulinum toxin treatment in spasmodic dysphonia: An H₂¹⁵O PET study. *Journal of Speech, Language, and Hearing Research, 49*(5), 1127–1146. [https://doi.org/10.1044/1092-4388\(2006/081\)](https://doi.org/10.1044/1092-4388(2006/081))
- Ballard, K. J., Halaki, M., Sowman, P., Kha, A., Daliri, A., Robin, D. A., Tourville, J. A., & Guenther, F. H. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Frontiers in Human Neuroscience, 12*, 510. <https://doi.org/10.3389/fnhum.2018.00510>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>
- Battistella, G., Fuertinger, S., Fleysler, L., Ozelius, L. J., & Simonyan, K. (2016). Cortical sensorimotor alterations classify clinical phenotype and putative genotype of spasmodic dysphonia. *European Journal of Neurology, 23*(10), 1517–1527. <https://doi.org/10.1111/ene.13067>
- Battistella, G., Kumar, V., & Simonyan, K. (2018). Connectivity profiles of the insular network for speech control in healthy individuals and patients with spasmodic dysphonia. *Brain Structure and Function, 223*(5), 2489–2498. <https://doi.org/10.1007/s00429-018-1644-y>
- Battistella, G., Termsarasab, P., Ramdhani, R. A., Fuertinger, S., & Simonyan, K. (2017). Isolated focal dystonia as a disorder of large-scale functional networks. *Cerebral Cortex, 27*(2), 1203–1215. <https://doi.org/10.1093/cercor/bhv313>
- Bianchi, S., Battistella, G., Huddleston, H., Scharf, R., Fleysler, L., Rumbach, A. F., Frucht, S. J., Blitzer, A., Ozelius, L. J., & Simonyan, K. (2017). Phenotype- and genotype-specific structural alterations in spasmodic dysphonia. *Movement Disorders, 32*(4), 560–568. <https://doi.org/10.1002/mds.26920>
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International, 5*(9/10), 341–345.
- Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*(4), 837–845. <https://doi.org/10.1093/cercor/bhm131>
- Christoffels, I. K., Van de ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex. *PLOS ONE, 6*(5), e18307. <https://doi.org/10.1371/journal.pone.0018307>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage, 9*(2), 179–194.
- Daliri, A., & Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *Journal of Neurophysiology, 122*(2), 552–562. <https://doi.org/10.1152/jn.00662.2018>
- Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex, 99*, 55–68. <https://doi.org/10.1016/j.cortex.2017.10.019>
- Erickson, M. L. (2003). Effects of voicing and syntactic complexity on sign expression in adductor spasmodic dysphonia. *American Journal of Speech-Language Pathology, 12*(4), 416–424. [https://doi.org/10.1044/1058-0360\(2003/087\)](https://doi.org/10.1044/1058-0360(2003/087))
- Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology, 104*(2), 1177–1194. <https://doi.org/10.1152/jn.00032.2010>
- Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics, 5*, 13. <https://doi.org/10.3389/fninf.2011.00013>
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*(5), 350–365. <https://doi.org/10.1016/j.jcomdis.2006.06.013>
- Guenther, F. H. (2016). *Neural control of speech*. MIT Press.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping, 7*(3), 213–223. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)7:3<213::AID-HBM5>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N)
- Haslinger, B., Erhard, P., Dresel, C., Castrop, F., Roettinger, M., & Ceballos-Baumann, A. O. (2005). “Silent event-related” fMRI reveals reduced sensorimotor activation in laryngeal dystonia. *Neurology, 65*(10), 1562–1569. <https://doi.org/10.1212/01.wnl.0000184478.59063.db>
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135–145. <https://doi.org/10.1038/nrn3158>
- Hirano, S., Kojima, H., Naito, Y., Tateya, I., Shoji, K., Kaneko, K. I., Inoue, M., Nishizawa, S., & Konishi, J. (2001). Cortical

- dysfunction of the supplementary motor area in a spasmodic dysphonia patient. *American Journal of Otolaryngology—Head and Neck Medicine and Surgery*, 22(3), 219–222. <https://doi.org/10.1053/ajot.2001.23436>
- Houde, J. F., & Nagarajan, S. S.** (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82. <https://doi.org/10.3389/fnhum.2011.00082>
- Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N.** (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364. <https://doi.org/10.1016/j.neuroimage.2012.02.055>
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E.** (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124–132. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Kirke, D. N., Battistella, G., Kumar, V., Rubien-Thomas, E., Choy, M., Rumbach, A. F., & Simonyan, K.** (2017). Neural correlates of dystonic tremor: A multimodal study of voice tremor in spasmodic dysphonia. *Brain Imaging and Behavior*, 11(1), 166–175. <https://doi.org/10.1007/s11682-016-9513-x>
- Kiyuna, A., Kise, N., Hiratsuka, M., Kondo, S., Uehara, T., Maeda, H., Ganaha, A., & Suzuki, M.** (2017). Brain activity in patients with adductor spasmodic dysphonia detected by functional magnetic resonance imaging. *Journal of Voice*, 31(3), 379.e1–379.e11. <https://doi.org/10.1016/j.jvoice.2016.09.018>
- Kiyuna, A., Maeda, H., Higa, A., Shingaki, K., Uehara, T., & Suzuki, M.** (2014). Brain activity related to phonation in young patients with adductor spasmodic dysphonia. *Auris Nasus Larynx*, 41(3), 278–284. <https://doi.org/10.1016/j.anl.2013.10.017>
- Kleber, B., Friberg, A., Zeitouni, A., & Zatorre, R.** (2017). Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage*, 147, 97–110. <https://doi.org/10.1016/j.neuroimage.2016.11.059>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lhéricy, S., Tijssen, M. A. J., Vidailhet, M., Kaji, R., & Meunier, S.** (2013). The anatomical basis of dystonia: Current view using neuroimaging. *Movement Disorders*, 28(7), 944–957. <https://doi.org/10.1002/mds.25527>
- Lenth, R.** (2019). *emmeans: Estimated marginal means, aka least-squares means*. R package. <https://cran.r-project.org/package=emmeans>
- Liu, H., & Larson, C. R.** (2007). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *The Journal of the Acoustical Society of America*, 122(6), 3671–3677. <https://doi.org/10.1121/1.2800254>
- Ludlow, C. L.** (2011). Spasmodic dysphonia: A laryngeal control disorder specific to speech. *Journal of Neuroscience*, 31(3), 793–797. <https://doi.org/10.1523/JNEUROSCI.2758-10.2011>
- MacDonald, E. N., Goldberg, R., & Munhall, K. G.** (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, 127(2), 1059–1068. <https://doi.org/10.1121/1.3278606>
- Max, L., & Maffett, D. G.** (2015). Feedback delays eliminate auditory–motor learning in speech production. *Neuroscience Letters*, 591, 25–29. <https://doi.org/10.1016/j.neulet.2015.02.012>
- Mitsuya, T., Munhall, K. G., & Purcell, D. W.** (2017). Modulation of auditory–motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America*, 141(4), 2758–2767. <https://doi.org/10.1121/1.4981139>
- Nieto-Castañón, A., & Fedorenko, E.** (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, 63(3), 1646–1669. <https://doi.org/10.1016/j.neuroimage.2012.06.065>
- Oldfield, R. C.** (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Olthoff, A., Baudewig, J., Kruse, E., & Dechent, P.** (2008). Cortical sensorimotor control in vocalization: A functional magnetic resonance imaging study. *The Laryngoscope*, 118(11), 2091–2096. <https://doi.org/10.1097/MLG.0b013e31817fd40f>
- R Core Team.** (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Revelle, W.** (2018). *psych: Procedures for personality and psychological research*. Northwestern University. <https://cran.r-project.org/package=psych>
- Roy, N., Gouse, M., Mauszycki, S. C., Merrill, R. M., & Smith, M. E.** (2005). Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *The Laryngoscope*, 115(2), 311–316. <https://doi.org/10.1097/01.mlg.0000154739.48314.ee>
- Schweinfurth, J. M., Billante, M., & Courey, M. S.** (2002). Risk factors and demographics in patients with spasmodic dysphonia. *The Laryngoscope*, 112(2), 220–223. <https://doi.org/10.1097/00005537-200202000-00004>
- Simonyan, K.** (2014). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurobiology*, 28, 15–21. <https://doi.org/10.1016/j.conb.2014.05.006>
- Simonyan, K., Berman, B. D., Herscovitch, P., & Hallett, M.** (2013). Abnormal striatal dopaminergic neurotransmission during rest and task production in spasmodic dysphonia. *Journal of Neuroscience*, 33(37), 14705–14714. <https://doi.org/10.1523/JNEUROSCI.0407-13.2013>
- Simonyan, K., & Horwitz, B.** (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, 17(2), 197–208. <https://doi.org/10.1177/1073858410386727>
- Simonyan, K., & Ludlow, C. L.** (2010). Abnormal activation of the primary somatosensory cortex in spasmodic dysphonia: An fMRI study. *Cerebral Cortex*, 20(11), 2749–2759. <https://doi.org/10.1093/cercor/bhq023>
- Simonyan, K., & Ludlow, C. L.** (2012). Abnormal structure–function relationship in spasmodic dysphonia. *Cerebral Cortex*, 22(2), 417–425. <https://doi.org/10.1093/cercor/bhr120>
- Simonyan, K., Tovar-Moll, F., Ostuni, J., Hallett, M., Kalasinsky, V. F., Lewin-Smith, M. R., Rushin, E. J., Vortmeyer, A. O., & Ludlow, C. L.** (2008). Focal white matter changes in spasmodic dysphonia: A combined diffusion tensor imaging and neuropathological study. *Brain*, 131(2), 447–459. <https://doi.org/10.1093/brain/awm303>
- Terumitsu, M., Fujii, Y., Suzuki, K., Kwee, I. L., & Nakada, T.** (2006). Human primary motor cortex shows hemispheric specialization for speech. *NeuroReport*, 17(11), 1091–1095. <https://doi.org/10.1097/01.wnr.0000224778.97399.c4>
- Tourville, J. A., & Guenther, F. H.** (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>
- Whitfield-Gabrieli, S., & Nieto-Castañón, A.** (2012). CONN: A functional connectivity toolbox for correlated and anti-correlated brain networks. *Brain Connectivity*, 2(3), 125–141. <https://doi.org/10.1089/brain.2012.0073>
- Zoons, E., Booij, J., Nederveen, A. J., Dijk, J. M., & Tijssen, M. A. J.** (2011). Structural, functional and molecular imaging of the brain in primary focal dystonia—A review. *NeuroImage*, 56(3), 1011–1020. <https://doi.org/10.1016/j.neuroimage.2011.02.045>