Effects of acoustic and articulatory perturbation on cortical activity during speech production

Jason Tourville[1], Frank Guenther[1,2,3], Satrajit Ghosh[2], Kevin Reilly[1], Jason Bohland[1], Alfonso Nieto-Castanon[1]

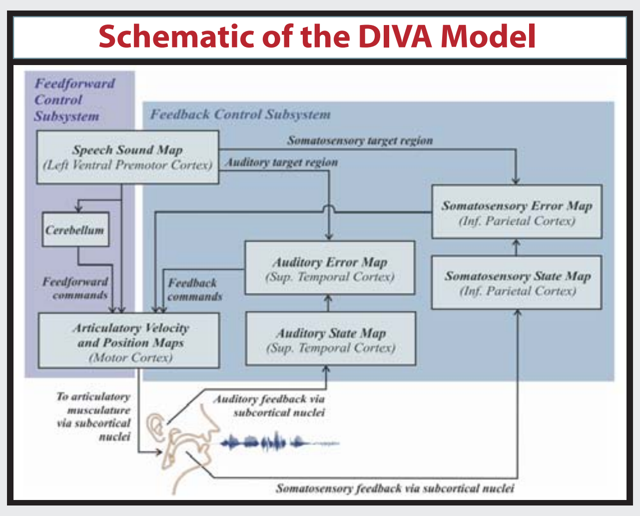
[1]Department of Cognitive and Neural Systems, Boston University, [2]Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, [3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital

Background

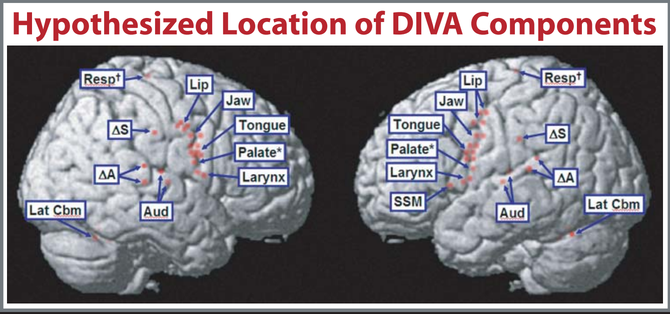
A number of studies have demonstrated rapid compensation for perturbation of speakers' articulator movements [1,2] and auditory feedback [3,4]. In the DIVA model of speech production [5] these compensatory behaviors result from a mismatch between expected and actual sensory feedback computed in higher-order sensory areas. Specifically, the model predicts increased activation in supramarginal gyrus due to articulatory perturbation and increased activation in posterior superior temporal gyrus and planum temporal during altered auditory feedback.

To test these predictions, we performed two fMRI experiments. In the first experiment we perturbed subjects' jaw movements unexpectedly during speech production. In the second, we unexpectedly perturbed subjects' auditory feedback.

Schematic of the DIVA Model



Hypothesized Location of DIVA Components



Methods

Experiment 1

Study Design

In each trial, a 2-syllable pseudo-word or control stimulus was presented:

Test stimuli: [anu, ani, agu, agi, atu, ati, au, ai]

Control: [yyy]

Subject instructed to:

- read word aloud as soon as it appears on screen
- remain silent during control trial
- keep mouth closed when not speaking

During testing, the subject's jaw was perturbed by the inflation of a balloon on 1/7 of pseudo-word trials and 1/2 of control trials.

- Perturbation triggered by subject vocalization at start of word, causing balloon to inflate while the jaw was lowered for the initial vowel "a"

4 Conditions:
1) No Perturbation, No Speech (baseline)
2) Unperturbed Speech
3) Perturbation Alone
4) Perturbed Speech

Each run had 7 of each pseudo-word + 16 controls = 72 trials per run, in random order

Stimulus duration = 3 s; Perturbation duration = 1.8 s; Trial length = 11 s.

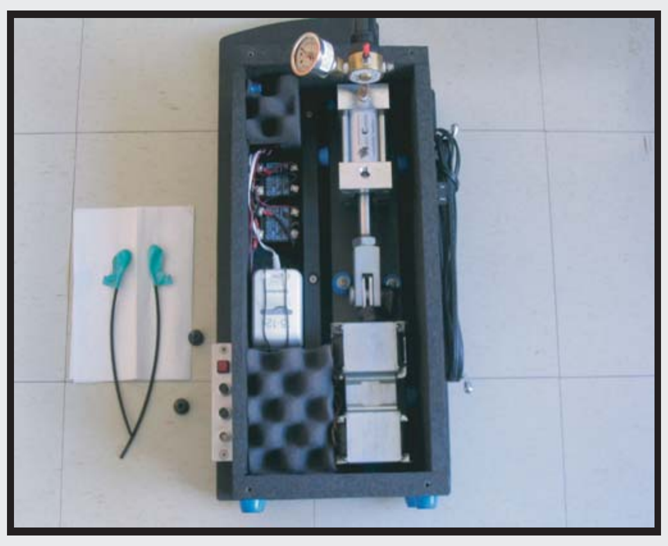
Subjects performed 2-5 runs (each run approx. 13 minutes)

13 subjects participated (6 female, 7 male)

Jaw Perturbation

A solenoid-driven air cylinder (a.k.a. the "Perturbatron") delivered pneumatic pressure (4 psi) to a small balloon (constructed from a heavy-duty rubber glove finger) placed between the subject's molars to perturb the jaw. The balloon fully inflates within 100 ms.

Perturbatron: pneumatic jaw perturbation device



Experiment 2

Study Design

In each trial, one of 8 words or a control stimulus was presented:

Test stimuli: [beck, bet, deck, debt, peck, pep, ted, tech]

Control: [yyy]

Subject instructed to:

- read word aloud as soon as it appears on screen
- remain silent during control trial

On 1 of every 4 test trials, the first formant of the subject vocalization was shifted up or down by 30%. The upward shift moved the vowel sound toward /ae/ the downward shift toward /I/.

4 Conditions:
1) No Shift (with vocalization)
2) Shift Up (F1*1.3)
3) Shift Down (F1*.7)
4) Control (no vocalization)

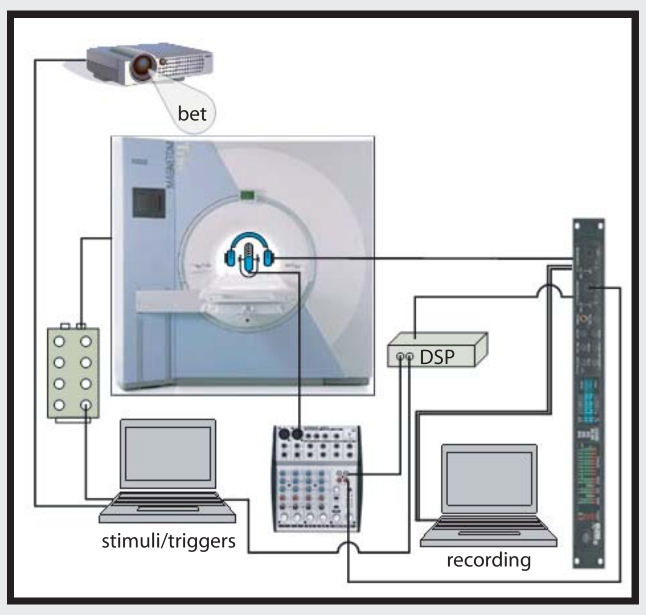
Each run consisted of 8 of each of the test words and 16 control stimuli = 80 trials/run (16 minutes per run)

Stimulus duration = 2 seconds, trial length = 12 seconds

11 subjects participated (5 female, 6 male)

Formant Shift

Subject vocalization received by Shure SM93 Micro-Lavelier microphone and fed back to the subject via Koss EXP-900 electrostatic headphones. Formant shifts were made by a TI DSK6713 DSP board (~17 ms delay).



Imaging

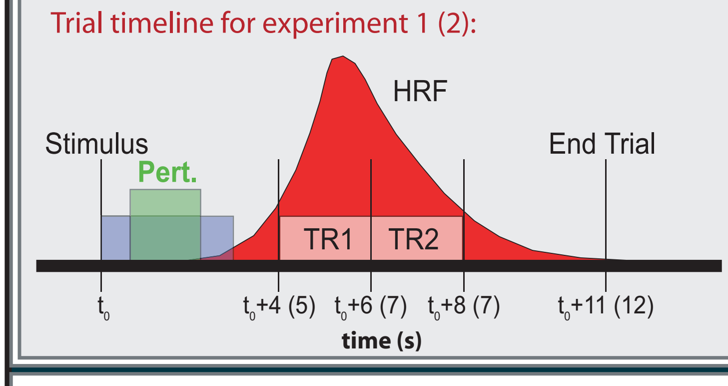
Stimulus presentation and perturbation/shift triggering performed with *Presentation Version 0.80* stimulus delivery software (www.neurobs.com).

Neurologically normal right handed English speakers participated. All subjects were recruited and provided informed consent in accordance with Boston University and Massachusetts General Hospital IRB guidelines.

Imaging was performed with a Siemens 3T Trio scanner at the Athinoula A. Martinos Center for Biomedical Imaging.

Event triggered design: 2 volumes were acquired per trial (TR = 2 s, 32 slices, 5 mm thickness) to minimize susceptibility artifact and permit vocalization in the absence of scanner noise.

Trial timeline for experiment 1 (2):



Data Analysis

Analysis was performed using the CNS SpeechLab Toolkit (http://speechlab.bu.edu), which integrates SPM2 [4], a region of interest (ROI) toolbox [5], and FreeSurfer [6] methods into a common object-oriented Matlab-based processing stream.

Two types of analyses were performed:

Voxel Based Analysis: Functional images were realigned, coregistered with a structural dataset, normalized into stereotaxic space, smoothed (12 mm FWHM Gaussian) and analyzed using SPM2. Data were analyzed using random-effects analysis, unless otherwise noted, and controlled for false discovery rate (FDR, denoted p_{FDR}) rate.

ROI Based Analysis: Freesurfer was used to create accurate cortical surface reconstructions from structural scans and to automatically parcellate [7] the surface according to a custom parcellation system for speech studies based upon the method developed at the Center for Morphometric Analysis at MGH [8]. Data were smoothed using a 12mm FWHM Gaussian kernel within ROIs and analyzed using random-effects analysis.

References

1. Abbs, J.H., and Gracco, V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, 51:705-723.

2. Kelso, J. A., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*.10(6):812-832.

3. Houde, J.F., Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*. 279(5354):1213-1216.

4. Xu, Y., Larson, C.R., Bauer, J.J., Hain, T.C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustic Society of America*. 116(2):1168-78.

3.Guenther, F.H. and Ghosh, S.S. (2003). A model of cortical and cerebellar function in speech. Proceedings of the XVth International Congress of Phonetic Sciences (pp. 169-173). Barcelona: 15th ICPhS Organizing Committee.

4. Friston K.J., Ashburner J., Holmes A., and Poline J. (2002) Statistical Parameteric Mapping. [www.fil.ion.ucl.ac.uk/spm/]

5. Nieto-Castanon, A., Ghosh, S., Tourville, J., and Guenther, F.H. (2003). Region of interest based analysis of functional imaging data. *Neuroimage*, 19(4): 1303-1316.

6. Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*. 9(2):195-207.

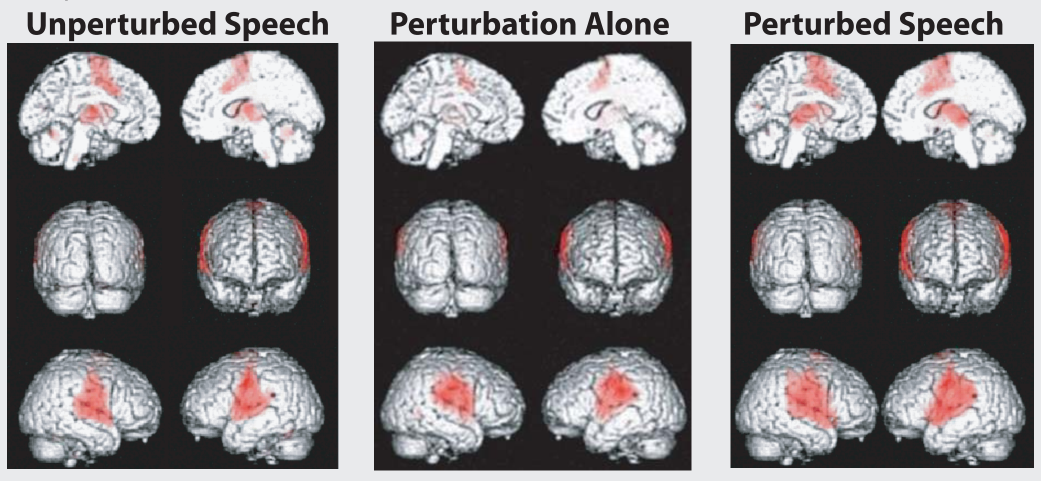
7. Fischl, B. et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*. 14:11-22.

8. Caviness, V.S., Meyer, J., Makris, N., Kennedy, D.N. (1996). MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience*. 8:566-587.

Articulatory Perturbation Results

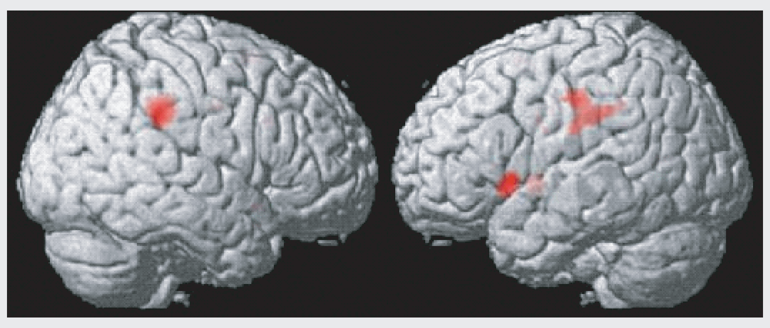
Baseline Contrasts ($p_{FDR} < .05$)

Comparison of the three test conditions vs. the baseline condition:



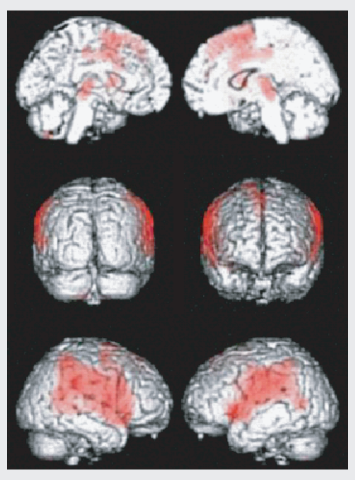
Unperturbed speech activated the expected sensorimotor, superior temporal, and medial frontal cortex as well as cerebellar and mid-brain regions. In addition to sensorimotor cortex activation, perturbation alone unexpectedly activated superior temporal and medial frontal cortex. Activation during perturbed speech expands posteriorly onto the supramarginal gyrus (SMg) and anteriorly into ventral premotor cortex (vPMC).

Perturbed vs. Unperturbed Speech $p_{FDR} < .001$ (masked by baseline contrasts)



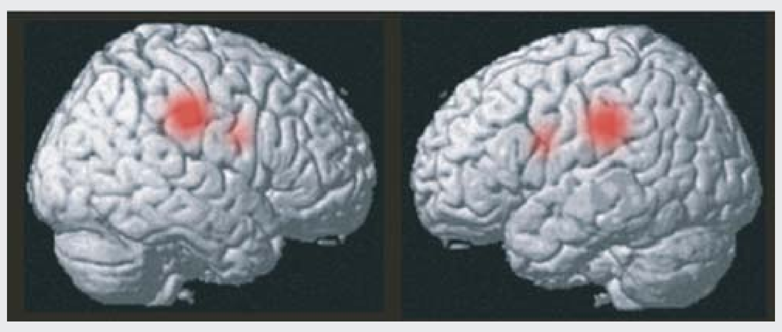
At a $p_{FDR} < 0.001$ threshold, bilateral SMg (MNI coordinates of local significance maximum: [-64, -38, 36] and [70, -32, 36] for left and right hemisphere, respectively) and left anterior insula (aINS, [-44 16 -4]) activation is seen. This can be compared to the location of simulated activations from the DIVA model (DIVA cell activations convolved with a hemodynamic response function):

Perturbed vs. Unperturbed Speech $p_{FDR} < .05$



With a $p_{FDR} < 0.05$ threshold, perturbed speech compared to unperturbed speech activates a wide expanse of cortex, extending to posterior SMg, posterior superior and middle temporal gyri, and inferior frontal cortex. At the right, activations are shown at a stricter threshold to highlight those regions most affected by the addition of the perturbation.

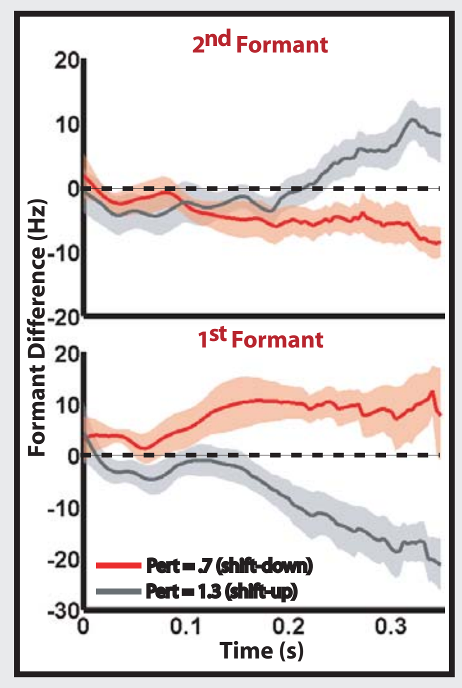
Regions *Predicted* to Respond Due to Perturbation



Bilateral SMg activation as a result of articulator perturbation is consistent with *a priori* hypotheses. Peak activation in aINS was not predicted, however.

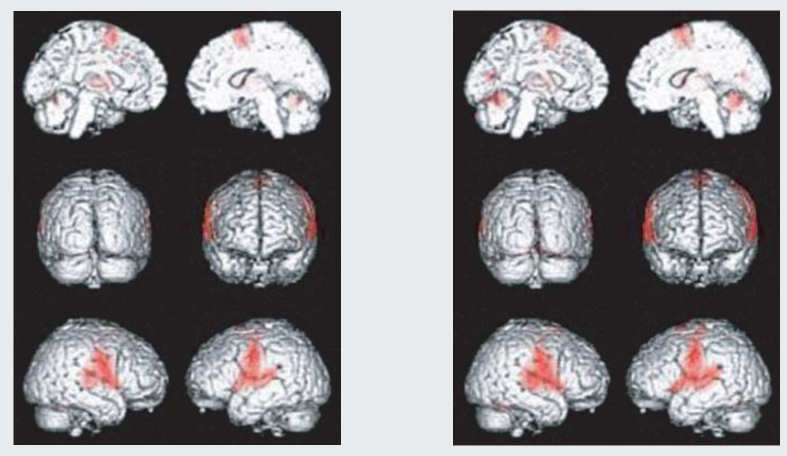
Auditory Shift Results

Subject Responses Controlled for Individual Formant Differences



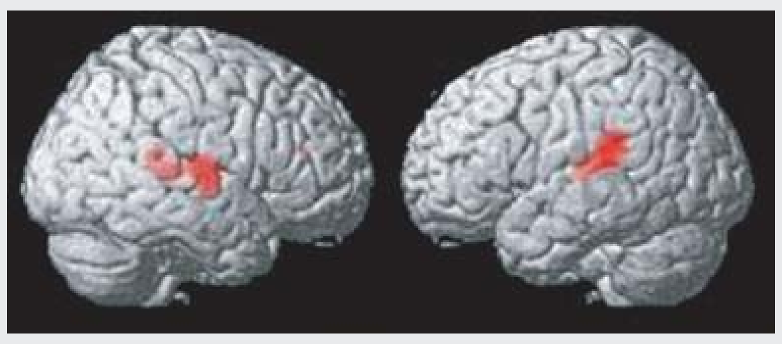
For each subject, an average contour was derived for the first two formants for the unshifted trials. To determine whether subjects compensated for the acoustic perturbation, the unshifted averages were subtracted from the formant contours produced during shifted trials and average shift-up and shift-down difference contours were derived. The figure on the left shows these contours, averaged across subjects, for F1 and F2. Compensation for the shift can be seen by the divergence of the red (downward shift) and gray (upward shift) lines from the dashed line as the trial proceeds.

Baseline Contrasts ($p_{FDR} < .05$)



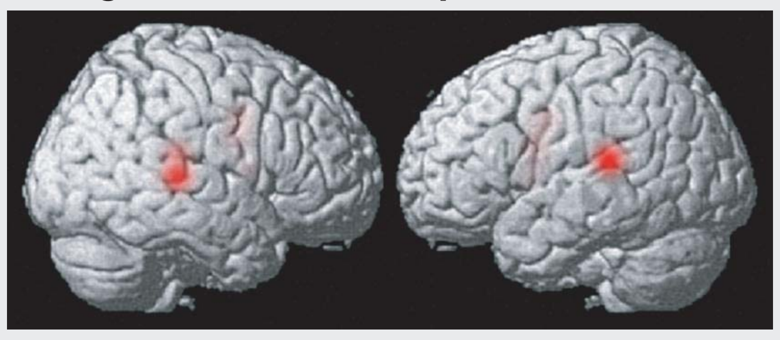
Unshifted speech activated the expected sensorimotor, superior temporal, and medial frontal cortex, cerebellar, and mid-brain regions when compared to baseline (*cf.* unperturbed vs. baseline contrast above). Shifted speech activated a similar network of regions with greater activation apparent in medial cerebellum, superior temporal gyrus, and medial frontal cortex.

Shift - No Shift Speech (fixed effects, $p_{FDR} < .05$) Masked by No Shift - Baseline contrast



Greater activation seen in: bilateral posterior superior temporal gyrus (pSTg, right [72, -32, 8], left [-58, -34, 20], the latter maximum located within left planum temporale (PT), bilateral central portion of pSTg (right [68, -16, 8], left [-58, -24, 10]), a region in right inferior frontal sulcus (IFs, [56, 40, 20]).

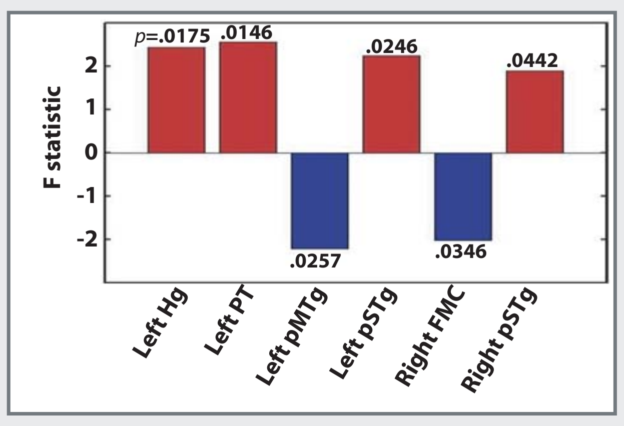
Regions *Predicted* to Respond Due to Shift



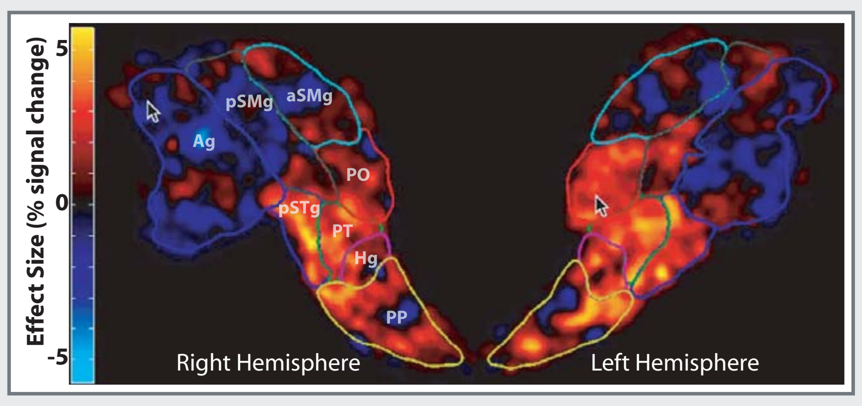
Bilateral pSTg and left PT activation as a result of auditory feedback shift is consistent with *a priori* hypotheses. Not predicted is the spread of superior temporal activation toward secondary and primary auditory cortex nor the right IFs activity. None of the differences seen in the contrast above survive random effects analysis corrected for multiple tests, however. Therefore, random effects analysis within regions of interest was performed. Results are shown in the box at right.

ROI Analysis of Shift - No Shift Contrast

Regions showing significant difference (random effects analysis)



Signal Change in Superior Temporal and Surrounding Areas



Ag = angular gyrus
aSMg = anterior supramarginal gyrus
FMC = frontomedial cortex
Hg = Heschl's gyrus
PP = planum polare
PO = parietal operculum
PT = planum temporale
pMTg = posterior middle temporal gyrus
pSMg = posterior supramarginal gyrus
pSTg = posterior superior temporal gyrus

Signal peaks in left pSTg (two peaks), PT, and Hg, and right pSTg, resulting in the significant region of interest based findings noted in the table above.

Conclusions

I. Perturbation of both articulatory and auditory feedback resulted in a general increase in activation of the speech production network relative to unperturbed speech.

II. Compared to unperturbed speech, speech during articulatory perturbation resulted in peak activations in/near bilateral supramarginal gyrus and left anterior insula.

III. Compared to unshifted speech, speech during shifted auditory feedback resulted in greater activation of posterior superior temporal gyrus bilaterally, and left Heschl's gyrus and planum temporale (these regions confirmed by ROI analysis).

IV. The results are consistent with the DIVA model predictions that, i) somatosensory error cells, coding the difference between expected and actual somatosensory feedback, lie in supramarginal gyrus, and ii) auditory error cells, coding the difference between expected and actual auditory feedback, lie in the posterior superior temporal gyrus and planum temporale.

V. In both experiments, significant activation due to perturbation spread toward primary sensory cortices more than expected, supporting the involvement of lower order sensory cortex in the processing of speech feedback error in addition to the higher order sensory cortical areas.

VI. In both experiments, anterior regions of activation were found that are thus far not accounted for by the DIVA model. Articulatory perturbation caused peak activation in left anterior insula while formant shifting led to activation of right inferior frontal sulcus (the latter did not survive FDR correction of random effects results, however).

VII. To the best our knowledge this is the first demonstration of short-latency compensation to formant perturbation.

This study was supported by NIH grant R01DC02852 (P.I. Frank Guenther). Use of the Athinoula A. Martinos Center for Biomedical Imaging facilities was aided by support from National Center for Research Resources grant P41RR14075 and the MIND Institute. We would like to thank Lawrence Wald, Bruce Fischl and Julie Goodman of the Martinos Center for their assistance.

Contact: http://speechlab.bu.edu; Jason Tourville: jtour@cns.bu.edu.

THE MIND INSTITUTE