INFLUENCE OF BRAIN REGIONS INVOLVED WITH ARTICULATORY PROCESSING ON PHONEME IDENTIFICATION PERFORMANCE

Daniel Callan[1,2] and Mitsuo Kawato[1]

[1]ATR Computational Neuroscience Laboratories, Kyoto Japan,
[2]National Institute of Communication Technologies, Kyoto Japan
dcallan@atr.jp kawato@atr.jp

ABSTRACT

This study investigates neural processes related to phoneme identification in the presence of white noise. Differential brain activity for a difficult consonant identification task (/b/ versus /d/) relative to an easier vowel identification task (/a/ versus /o/) using identical stimuli was present in brain regions involved with speech articulatory planning control (Broca's area, anterior insula, premotor cortex), instantiation of internal models (cerebellum), and auditory processing regions (STG/S). The results of a correlation analysis of behavioral performance with brain activity, as well as an analysis of incorrect versus correct responses suggests that activity in brain regions involved with articulatory planning control is related with poorer performance. These results are inconsistent with hypotheses that articulatory planning areas (Broca's area, anterior insula, premotor cortex) are utilized to facilitate speech perception. Considerable activity in the cerebellum for correct relative to incorrect responses is consistent with the hypothesis that articulatory-auditory internal models instantiated in the cerebellum are utilized to facilitate phoneme perceptual identification performance.

Keywords: fMRI, Internal-Model, Perception, Broca's, Cerebellum.

1. INTRODUCTION

The use of articulatory features of speech production to mediate and/or constrain speech perception has been proposed long ago [4,6]. These ideas have recently been reformulated in the proposed 'mirror neuron theory' [5] and the concept of 'internal models' [1]. The 'mirror neuron theory' essentially maintains that perception is mediated by processes in brain regions involved with production of the observed action. The 'internal-models' position is essentially a reformulation of 'analysis-by-synthesis' [6] in that it maintains that speech perception is facilitated by constraints of a model of the articulatory – auditory mapping of the speech production system. The primary impetus for proposing the use of articulatory constraints for perception was the apparent lack of invariant features defining phonemes in the acoustic signal. However, this is an issue of some debate and there are a number of researchers that maintain that there is sufficient information in the acoustic signal to specify phonetic identification without imposing articulatory constraints.

It has been proposed that under conditions when the auditory signal is clear and the phonemes are not ambiguous (e.g. native phonetic contrasts) brain regions involved with acoustic phonetic processing of aspects of the auditory signal (anterior superior temporal gyrus/sulcus STG/S) are predominantly used for speech perception. However, under conditions when the auditory signal is unclear and/or the sounds are ambiguous (e.g. non-native phonetic contrasts) brain regions involved with speech articulation planning (Broca's area, premotor cortex, anterior insula) are used to facilitate speech perception. The auditory consequences of internally simulated articulatory control signals (articulatory-auditory internal models for various phonemes) are used to constrain speech perception based on competitive selection of the control signal that best matches the ongoing auditory signal. Recurrent mappings between brain regions involved with speech articulation planning and those involved with audiory gestural mapping (planum temporale, posterior STG) instantiate these articulatory-auditory internal models. Evidence suggests that the cerebellum also instantiates 'internal models' [2,3]. The cerebellum potentially may form internal models of the cortical dynamics of the interactions of several brain regions involved with articulatory-auditory mapping and may serve as an update signal for smoother more efficient processing.

The experiment presented here utilizes a speech in noise task to investigate neural processes involved with phoneme perception. The task involves identifying whether the syllable presented begins with /b/ or /d/. Previous research supporting the use of internal models has shown performance related increases in brain activity in regions involved with speech articulation planning as well as the cerebellum for perceptual identification of a non-native contrast [1]. It is predicted that if indeed articulatory features are used to constrain and facilitate auditory phonetic perceptual identification that there will be performance related increases in activity in brain regions involved with instantiating the articulatory-auditory internal models. Alternatively, if auditory feature extraction processes are primarily utilized to facilitate phonetic perception in the presence of noise then performance related increases in activity are predicted to occur in auditory processing brain regions (STG/S).

2. METHODS

2.1. Subjects

Thirteen right-handed individuals (12 male, one female) ranging from 25 to 42 years of age participated in this study. The phonemes used in the experiment fell within native categories for all the subjects. All subjects gave written informed consent for experimental procedures approved by the ATR Human Subject Review Committee.

2.2. Stimuli and Procedure

The stimuli consisted of the following synthesized speech sounds /ba/, /bo/, /da/, /do/ that were band passed filtered from 300 to 3400 Hz. Each of the stimuli was 120 msec in duration and normalized to have the same rms energy. White noise was constructed and band passed filtered from 300 to 3400 Hz. All sound files were sampled at 44100 Hz. In order to account for greater perceived loudness for /a/ stimuli over /o/ stimuli the /ba/ and /da/ stimuli were presented at a 1 dB signal-to-noise ratio and the /bo/ and /do/ stimuli were presented at a 4 dB signal-to-noise ratio.

The task was two-alternative forced-choice phoneme identification in the presence of white noise. Subjects were asked to respond quickly while maintaining accuracy. There were three conditions: 1. Consonant identification (/b/ versus /d/); 2. Vowel identification (/a/ versus /o/); 3. Passive listening to only white noise. A blocked design was used in which 6 trials were presented (the instructions as to the task condition were given visually at the start of each block). The /bd/ condition was presented after each subsequent block of either /ao/ or /passive/ trials. There were 8 blocks in each run and a total of 8 runs for the entire fMRI experiment. The order of the blocks was counterbalanced across runs. For each trial the noise before stimulus presentation was present for 1000-2500 msec, after the stimulus was presented the noise remained on for 1500 msec, the time between trials was between 1500-2000 msec. The block duration was fixed at 38 seconds.

2.3. Data Collection and Analysis

An event-related procedure was utilized to analyze the fMRI data even though it was collected using a blocked design. The event-related analysis procedure allows for focusing on brain activity from the onset of the speech stimuli, whereas, a blocked analysis would reflect to a greater extent activity related to noise prior to the speech stimulus. The stimuli were presented using Neurobehavioral System's Presentation software via MR-compatible headphones. For functional brain imaging, Shimadzu-Marconi's Magnex Eclipse 1.5T PD250 was used at the ATR Brain Activity Imaging Center. Functional T2* weighted images were acquired using a gradient echo-planar imaging sequence (echo time 55ms; repetition time 2000ms; flip angle 90°). A total of 20 sequential axial slices were acquired with a 3.5x3.5x6mm voxel resolution (one mm gap) covering the cortex and cerebellum. A total of 152 scans were taken for a single session. Images were preprocessed using programs within SPM2 (Wellcome Department of Cognitive Neurology, UCL). Differences in acquisition time between slices were accounted for, images were realigned and spatially normalized to a standard space using a template EPI image (3x3x3 mm voxels), and were smoothed using a 7x7x12 mm FWHM Gaussian kernel. Regional brain activity for the various conditions was assessed using a general linear model employing a boxcar function convolved with a hemodynamic response function. Random effects analysis was used for all contrasts of interest (consonant task versus passive; vowel task versus passive; consonant task versus vowel) thresholded at $p < 0.005$. An additional event-related design was utilized to extract the same number of correct and incorrect trials for the consonant task within each run separately for each subject. In this way brain activity related to correct versus incorrect responses for the consonant task could be identified (random effects analysis; threshold $p < 0.005$).

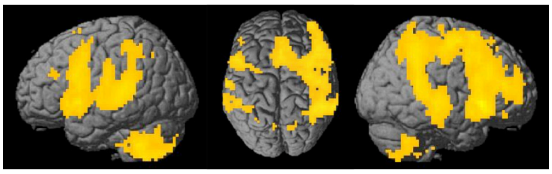
3. RESULTS

Behavioral performance for the consonant identification task (mean 79.09%, std 9.64%) is significantly lower ($T = 5.4$; $p < 0.001$) than that of the vowel identification task (mean 96.16%, std 4.49%).

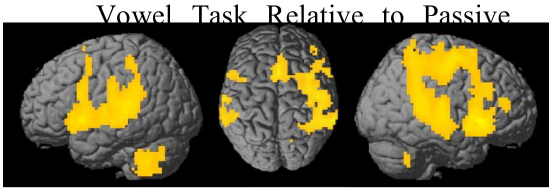
The fMRI results for the consonant task relative to the passive noise condition and the vowel task relative to the passive noise condition are presented in Fig. 1. For both contrasts brain activity is present bilaterally in auditory processing regions (STG/S), articulatory planning regions (Broca's area, premotor cortex, andterior insula), regions involved with orosensory processing (the supramarginal gyrus SMG) and brain regions thought to instantiate internal models (cerebellum).

Figure 1: Task related brain activity.

Consonant Task Relative to Passive

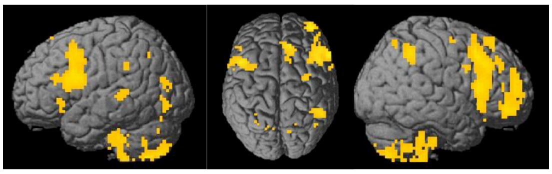


Vowel Task Relative to Passive



The contrast of the consonant task versus the vowel task (Fig. 2) focuses on processes related to extraction and identification of the phonemes for a difficult relative to an easy task. And controls for brain activity related to the acoustics of the stimuli as well as properties of the task such as the processes related to the button press decision. The consonant versus the vowel contrast (Fig. 2) shows considerable differential activity bilaterally in articulatory planning regions (Broca's area, premotor cortex, andterior insula), and brain regions thought to instantiate internal models (cerebellum) as well as some activity in the left STG/S.

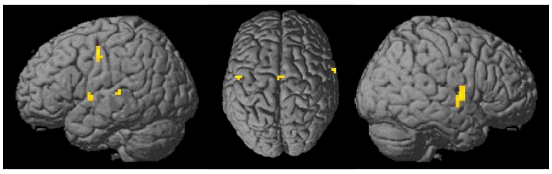
Figure 2: Consonant relative to vowel contrast.



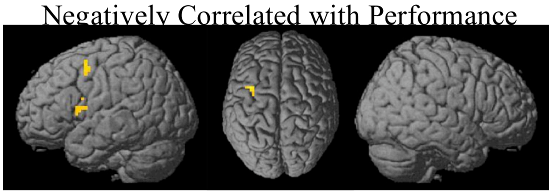
Correlation analysis of consonant task performance with that of brain activity for the consonant relative to vowel contrast (Fig. 3) reveals a positive relationship primarily in auditory processing regions, bilateral anterior STG, left STG, and left dorsal premotor cortex. In contrast a negative relationship with performance was identified in speech articulatory planning regions (left Broca's area, left anterior insula, and left premotor cortex).

Figure 3: Correlation of perceptual consonant identification performance and brain activity.

Positively Correlated with Performance



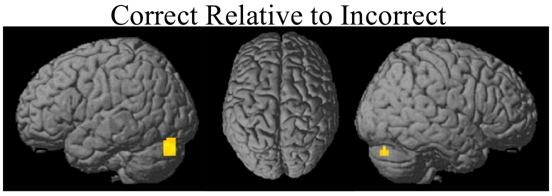
Negatively Correlated with Performance



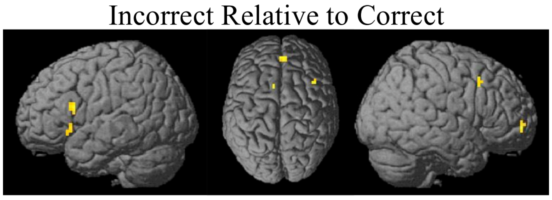
Analysis of correct versus incorrect responses for the consonant task reveals differential activity in the cerebellum bilaterally, whereas, the analysis of incorrect versus correct responses reveals

Figure 4: Brain activity related to correct and incorrect responses.

Correct Relative to Incorrect



Incorrect Relative to Correct



differential activity in speech articulatory planning regions (left Broca's area, left anterior insula, and right premotor cortex (Fig. 4).

4. DISCUSSION

The results indicate a similar pattern of activity involving speech articulatory planning areas as well as auditory processing areas for both consonant and vowel tasks relative to the passive listening condition (Fig.1). The considerable differential activity in brain regions thought to be involved with internal simulation of speech production (Broca's area, premotor cortex, anterior insula, and cerebellum) for the consonant task relative to the vowel task (Fig. 2) is consistent with the hypothesis that these regions are utilized to facilitate perceptual performance. Alternatively, the activity may rather just reflect the use of more sub-vocal rehearsal when phoneme identification is unclear that is not related to facilitating speech perception.

In order to investigate the relationship between task performance and brain activity involved with perceptual identification a correlation analysis was conducted. The analysis revealed a positive relationship between behavioral performance on the consonant task and brain activity in primarily left and right auditory processing regions (left and right anterior STG, and left posterior STG/S). A negative relationship with task performance is present in articulatory planning regions (left Broca's area, left anterior insula, and left premotor cortex). These results suggest that better subjects have more activity in auditory processing regions and less in articulatory planning regions than worse subjects. These results are in contrast to correlation between task performance and brain activity for difficult non-native phonetic identification in which brain activity in speech articulatory planning regions shows a positive relationship [1]. It is entirely possible that different processes are involved when there is a clear native auditory target for the phoneme to be identified in a speech in noise task versus when the auditory target is not well delineated (as in the case of non-native phoneme identification). The results of the correlation analysis lend support to the hypothesis that auditory processes are utilized to facilitate speech perception in the presence of noise and appear to falsify the hypothesis that articulatory planning regions are used to facilitate performance.

To explore within-subject relationships between performance and brain activity correct versus incorrect responses were analyzed for the consonant identification task. Correct relative to incorrect performance showed considerable differential activity in the cerebellum bilaterally (Fig. 4). Whereas, the incorrect relative to correct contrast showed differential activity in brain regions involved with speech articulatory planning (left Broca's area, left anterior insula, and right premotor cortex). Together with the correlation analysis above these results are inconsistent with the hypothesis that brain regions involved with articulatory planning control are used to facilitate performance. In contrast, the results suggest that activity in these regions hinders performance.

Given these pattern of results in brain regions involved with speech articulation planning one might dismiss the use of articulatory-auditory internal models to help facilitate perception. However, the considerable bilateral activity in the cerebellum for correct relative to incorrect responses (Fig.4) may indeed reflect greater use of articulatory-auditory internal models to facilitate performance. It is possible that well-established articulatory-auditory internal models instantiated in the cerebellum facilitate perception, whereas, internal simulation mediated in cortical regions interferes with cortical auditory perceptual processing.

5. REFERENCES

[1] Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage* 22, 1182-1194.

[2] Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., Yoshioka, T., Kawato, M., 2000. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature 403*, 192–195.

[3] Kawato, M., 1999. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol. 9*, 718–727.

[4] Liberman, A.M., Cooper, F.S., Shankweiler, D.S., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review 74*, 431-461.

[5] Skipper, J. Nusbaum, H., and Small, S., 2005. Lending a helping hand to hearing: Another motor theory of speech perception. In: *Action to Language Via the Mirror Neuron System* (Arbib, M. ed.). Cambridge, MA: Cambridge University Press.

[6] Stevens K. N., Halle M. 1967. Remarks on analysis by synthesis and distinctive features. In: *Models for the perception of speech and visual form* (Walthen-Dunn W, ed.). Cambridge, MA: MIT Press.