Brain Regions Involved with Speech Motor Control Mediate Phonetic Perceptual Identification Performance to a Greater Extent than Brain Regions Involved with Auditory Processing

Daniel Callan[1,2], Okito Yamashita[1], Noriko Yamagishi[1,2], Masaaki Sato[1], Mitsuo Kawato[1]

[1]ATR – Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan, 619-0288

[2]National Institute of Information and Communication Technology, Japan

dcallan@atr.jp, oyamashi@atr.jp, n.yamagishi@atr.jp sato@atr.jp kawato@atr.jp

Abstract. *This study investigates the relative influence that speech-motor versus auditory processing brain regions have on perceptual phonetic identification performance during degraded auditory conditions. An analysis procedure involving independent component analysis from single trial MEG data to extract spatial filters and determine classification features was used to assess the relative contribution of brain regions involved with speech motor control and those involved with auditory processing in predicting correct from incorrect perceptual identification. Additionally, classification of the presence of the perceptual identification task versus passive listening to white noise was conducted. The fMRI results showed greatest activity in brain regions involved with speech motor control for the speech identification task versus the passive listening task. However, the results of the classifier showed greater performance for the auditory processing areas than speech-motor areas. The primary result of this study indicates that classification of correct versus incorrect performance is mediated by speech-motor areas to a greater extent than auditory processing areas. The results are consistent with the hypothesis that auditory processing regions are responsible for picking up features in the auditory signal that are ambiguous and that phonetic identification performance is mediated by speech-motor areas that utilize articulatory–auditory mapping of speech production.*

1. Introduction

The potential role that speech production processes mediate speech perception has been proposed long ago (Stevens and Halle, 1967; Liberman et al., 1967). The main reason to invoke production constraints for speech perception was because of the lack of invariant features in the acoustic signal. More recently these ideas have been reformulated in the proposed 'mirror neuron theory' (Skipper et al., 2005) and the concept of 'internal models' (Callan et al., 2004). 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' (Stevens and Halle, 1967) in that it maintains that speech perception is facilitated by constraints of a model of the articulatory – auditory mapping of the speech production system.

Many brain imaging studies involving speech perception tasks have identified activity in brain regions involved with speech motor control as well as brain regions involved with auditory speech processing. The superior temporal gyrus/sulcus has been identified as a region selectively involved with auditory speech perception (Scott et al., 2000). The premotor cortex and Broca's area (brain regions thought to be involved with speech motor control) have been observed to be active during speech perception (Wilson et al., 2004) most especially when the task involves ambiguity (Callan et al., 2003). Greater activity in brain regions involved with auditory speech processing (superior/middle temporal gyrus) as well as brain regions involved with speech motor control (the PMC and Broca's area) have been associated with better performance in foreign language phonetic identification (Callan et al., 2003; 2004).

This study attempts to answer the question of whether phonetic identification under adverse auditory conditions (speech-in-noise task) utilizes brain regions involved with speech motor control (frontal regions including premotor cortex and Broca's area) to a greater extent than regions involved with extraction of auditory speech features (superior and middle temporal regions). According to theories proposing motor constraints for speech perception identification, performance should primarily be mediated by speech motor regions involved with 'analysis-by-synthesis'. As opposed to auditory regions that may be encoding multiple auditory features that do not unambiguously specify the speech signal. A novel approach to brain imaging analysis based on classification is used to address this question. Functional magnetic resonance imaging fMRI is used to show brain regions active during a two-alternative forced choice phonetic identification task in the presence of white noise over that of just passively listening to white noise. Magnetoencephalography MEG is used to acquire time courses of activity within frontal (speech motor regions) and temporal brain regions (auditory processing) for each trial that are analyzed using various techniques (wavelet analysis, entropy, root-mean-squared energy, peak-to-peak amplitude, phase). The results of which are used for training a classifier to distinguish between the phonetic identification task versus the passive listening task, as well as for correct versus incorrect performance on the phonetic identification task. It is hypothesized that both frontal and temporal regions will be able to classify the phonetic identification task versus the passive task but that frontal regions will be better at classifying correct versus incorrect performance for the phonetic identification task.

2. Methods

2.1. Subjects

One male native English-speaking subject 37 years of age participated in this study. The Participant was right-handed with no hearing or speech problems and 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 -4 signal-to-noise ratio and the /bo/ and /do/ stimuli were presented at a -1 signal-to-noise ratio during the MEG experiment;1 and 4 signal-to-noise ratio respectively was used for the fMRI experiment due to the additional noise created by the scanning procedure.

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. Place of articulation 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 16 runs for the entire experiment for MEG and 8 for fMRI. 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. In this study only the /bd/ and passive listening conditions are investigated.

2.3. Data Collection and Analysis: fMRI

The fMRI procedure consisted of a blocked design in which the auditory 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. The passive condition was implicitly modeled in the design. The F contrast for the /bd/ condition relative to the passive condition was assessed.

2.4. Data Collection and Analysis: MEG

The MEG procedure consisted of single-trial analysis in which the auditory stimuli were presented via insert earphones with a 3-meter tube leading outside the MEG room. For MEG, Yokogawa 208 channel system (laying down position) was used at the ATR Brain Activity Imaging Center. The channels were sampled at 1000 Hz and later down-sampled to 250 Hz. for analysis. Each of the trails was segmented from 500 msec before to 500 msec after stimulus onset. The same number of correct and incorrect trials for the /bd/ condition were randomly selected as well as a subset of trials from the /ao/ and passive conditions. These trials were then submitted to independent component analysis ICA (EEGLAB, Delorme and Makeig, 2004). ICA is an unsupervised method by which spatial filters can be constructed by unmixing activity at the sensors into maximally independent sources. ICA has the benefit of extracting artifacts (eye movement, blink, etc …) from the data. The spatial filters allow for activity within the specified region/s to be investigated across the entire time course even when the activity is very low. Two independent components with focal spatial filters over left and right frontal areas (likely reflecting Broca's and premotor cortex) and two independent components with focal spatial filters over left and right temporal areas (likely reflecting superior and middle temporal cortex) were selected for further analysis (See Figures 2-3 for spatial filters of each of the independent components selected).

In order to investigate differences in brain activity between the /bd/ and passive conditions as well as between correct and incorrect performance for the /bd/ condition a methodology employing classification was employed. This methodology reveals potential information recorded by MEG and analyzed by various methods that the brain may be utilizing for processing certain tasks. The ICA transformed data was subjected to multiple analyses. Gaussian wavelets (at 10, 14, 24, and 40 Hz) were used to generate amplitude and phase for each trial and time point (500msec post-stimulus down-sampled to 250 Hz = 125 points). In addition the trials were band-pass filtered at the following frequencies (low-alpha 8-12; high-alpha 12-16; beta 16-32; gamma 32-50) and then submitted to the following analysis methods: root-mean-squared RMS amplitude; Largest Peak-to-Peak amplitude; Entropy; Phase. All of the results from the different methods we will refer to as features. Only the activity after stimulus presentation was considered in this study. Eighty percent of the trials (bd vs passive: 246 total, 123 bd, 123 passive) (bd correct vs incorrect, 164 total; 82 correct, 82 incorrect) were used in selection of the features and to train the classifier. Twenty percent of the trials (bd vs passive: 62 total, 31 bd, 31 passive) (bd correct vs incorrect, 42 total, 21 correct, 21 incorrect) were used as a novel test validation set. The training features were normalized to mean zero and standard deviation of one. The normalize parameters from the training set were used to normalize the test set. Selection of the potential 1040 features for each independent component used to train the classifier was narrowed down based on an arbitrary statistical threshold for a t-test of the contrast of

interest (/bd/ vs. passive or correct vs. incorrect) to reduce the total number of entered features to be less than 150. Sparse logistic regression (Yamashita, et al., 2006) was used to classify the data. Sparse logistic regression utilizes a hierarchical Bayesian formulation to determine the relevance of the various features in an iterative manner. The relevance of many of the features is weighted at zero thus effectively removing these features from the analysis and serving as an automatic feature extraction process. In training of the classifier multiple iterations of cross validation were used to explore many combinations in the data set (80% train 20% test from original training set; 1000 cross-validation runs). Dominant features are based on the frequency of selection across cross- validation runs (Protects against over-fitting to one training data set). After features are selected via this cross-validation procedure the entire training set is used to adjust the weights of the sparse logistic regression classifier. The trained weights are then used to assess classification of the novel test trials.

3. Results

3.1. fMRI Results

Behavioral performance during fMRI for the /bd/ condition was 82.8% correct. The brain imaging results of the fMRI analysis for the contrast of the /bd/ condition relative to the passive listening condition are given in Figure 1. Activation is shown rendered on the surface of the brain for an F-test ($F = 23.85$, $p < 0.05$ corrected for multiple comparisons, with an extent threshold of 25 voxels). Activity is most prominent in the regions of Broca's area and the premotor cortex bilaterally. Activity is also present across the brain in the temporal lobes, paretial lobes, occipital cortex, and the cerebellum.

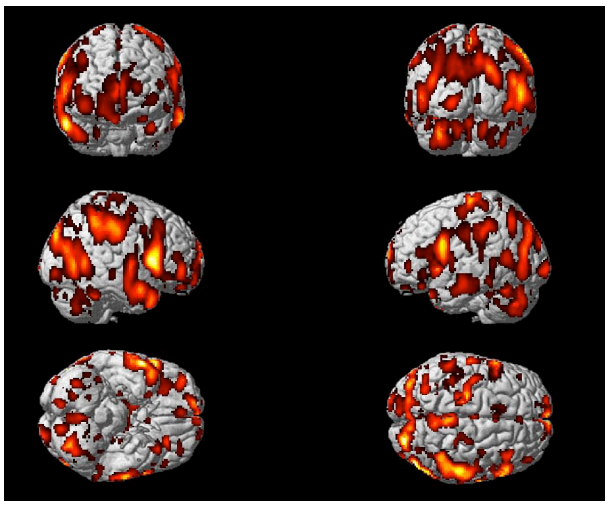

Figure 1. fMRI results for the contrast of the phonetic identification task relative to the passive listening task.

3.2. MEG Results

The behavioral performance for the /bd/ identification task during MEG was 66%. The mean activation waveforms for each of the independent components for the /bd/ versus passive and /bd/ correct versus incorrect are given in Figures 2-3. Each of the activation waveforms for the /bd/ condition shows a distinct event-related evoked potential (latency of peaks are given in Figures 2-3). Four different classification analyses using sparse-logistic regression were conducted: 1. Classification of /bd/ vs. passive conditions for the frontal independent components; 2. Classification of /bd/ vis passive conditions for the temporal independent components; 3. Classification of /bd/ correct vs. incorrect for the frontal independent components; 4. Classification of /bd/ correct vs. incorrect for the temporal independent components. For classification of /bd/ versus the passive condition, the temporal components outperformed the frontal components: temporal 75.8% classification relative to frontal 66.13% classification. However, for classification of /bd/ correct versus incorrect, the frontal components outperformed the temporal components: frontal 64.3% classification relative to temporal 40.5% classification.

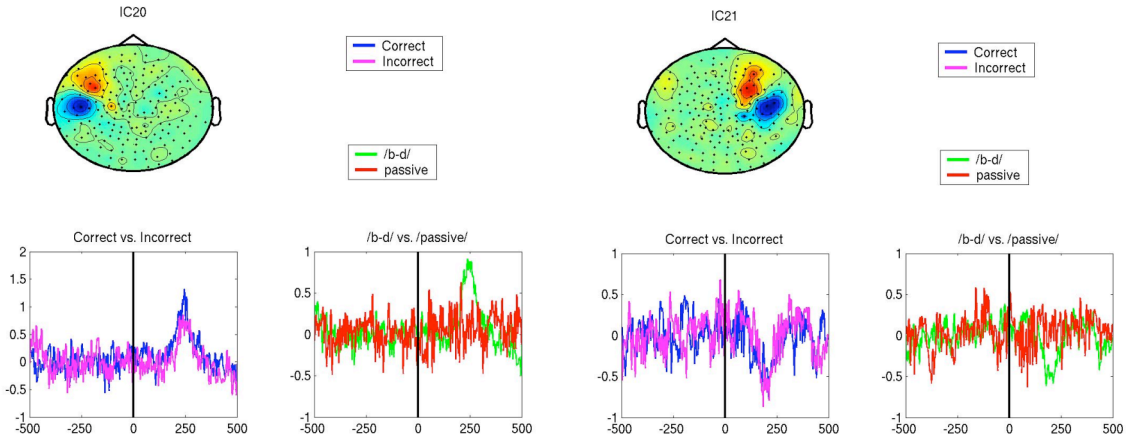

Figure 2. MEG results depicting spatial filter over left and right frontal areas respectively (top) and the plots of the mean activation waveform for the correct versus incorrect conditions and the /bd/ identification task versus the passive task (bottom).

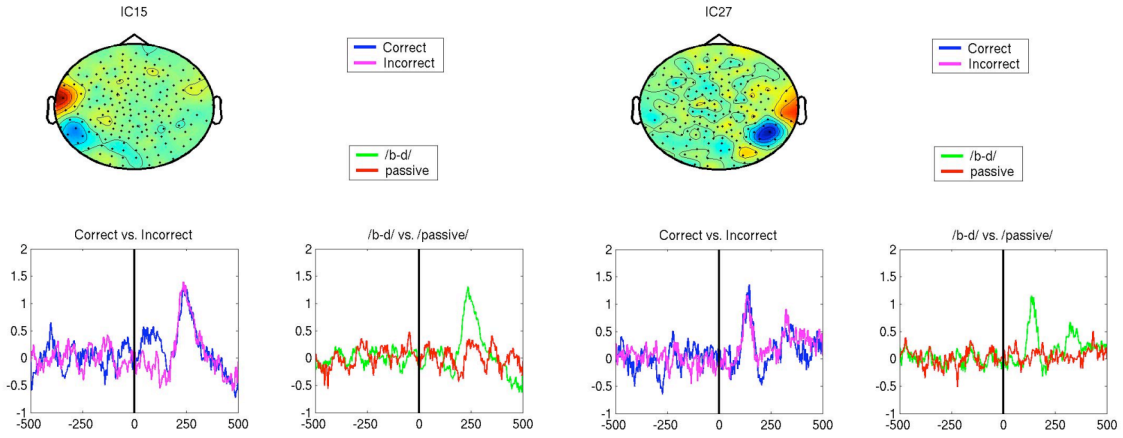

Figure 3. MEG results depicting spatial filter over left and right temporal areas respectively (top) and the plots of the mean activation waveform for the correct versus incorrect conditions and the /bd/ identification task versus the passive task (bottom).

4. Discussion

The results of this study support the hypothesis that brain regions involved with speech motor control mediate perceptual identification performance to a greater extent than brain regions involved with auditory processing. The classification performance for brain regions involved with speech motor control (frontal regions including Broca's area and the premotor cortex) have better classification performance than brain regions involved with auditory processing (temporal regions including superior and middle temporal gyrus) (see Figures 2-3). Additionally, the location of the greatest activity for the fMRI data for the phonetic identification versus the passive task is in the premotor cortex and Broca's area bilaterally (Figure 1). However, the classification results for the temporal regions are much better than those for frontal regions. These results are somewhat to be expected since there is an additional auditory stimulus being presented. It is likely that the temporal cortex activity in the fMRI (Figure 1) analyzed across blocks may be reduced when compared to the passive condition that also consists of considerable auditory activity from the presence of the white noise. Additionally, the fMRI analysis is across the entire block therefore, it contains activity prior as well as post stimulus presentation. The time resolution of MEG may allow for better extraction of brain activity corresponding to brief auditory signals than fMRI.

In summary, classification of the presence of a stimulus in a phonetic identification task is better classified by temporal regions involved with auditory processing. These regions may essentially be picking up features in the auditory signal that are ambiguous with regards to phoneme identification. Phonetic identification performance is mediated by frontal areas (Broca's area, premotor cortex). These results are completely consistent with the use of internal models that simulate the articulatory – auditory mapping of speech production in order to facilitate speech perception under degraded auditory conditions.

References

Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 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, 2004.

Callan, D.E., Tajima, K., Callan, A.M., Kubo, R., Masaki, S., Akahane-Yamada, R., Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19, 113-124, 2003.

Delorme, A. and Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134: 9-21, 2004.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., and Studdert-Kennedy, M. Perceptin of the speech code. *Psychological Review* 74: 431-461, 1967.

Scott, S. K., Blank, C. C., Rosen, S., Wise, R. J. S., Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400-2406, 2000.

Skipper, J. Nusbaum, H., and Small, S. 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, 2005.

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

Wilson, S. M., Sygin, A. P., Sereno, M. I., Iacoboni, M., Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7: 701-702, 2004.