



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Premotor cortex mediates perceptual performance

Daniel Callan^{a,*}, Akiko Callan^b, Mario Gamez^a, Masa-aki Sato^a, Mitsuo Kawato^a

^a ATR Computational Neuroscience Laboratories, Kyoto, Japan

^b NICT Multimodal Communication Group, Kyoto, Japan

ARTICLE INFO

Article history:

Received 13 November 2009

Revised 9 February 2010

Accepted 10 February 2010

Available online 22 February 2010

Keywords:

fMRI

MEG

Perception

Speech

Mirror Neuron

Motor theory

Premotor cortex

Broca's area

Event-related desynchronization

Event-related synchronization

Alpha

Gamma

Phonetic

ABSTRACT

Articulatory goals have long been proposed to mediate perception. Examples include direct realist and constructivist (analysis by synthesis) theories of speech perception. Although the activity in brain regions involved with action production has been shown to be present during action observation (Mirror Neuron System), the relationship of this activity to perceptual performance has not been clearly demonstrated at the event level. To this end we used functional magnetic resonance imaging fMRI and magnetoencephalography MEG to measure brain activity for correct and incorrect trials of an auditory phonetic identification in noise task. fMRI analysis revealed activity in the premotor cortex including the neighboring frontal opercular part of Broca's area (PMC/Broca's) for both perception and production tasks involving the same phonetic stimuli (potential mirror system site) that was significantly greater for correct over incorrect perceptual identification trials. Time–frequency analysis of single trials conducted over MEG current localized to PMC/Broca's using a hierarchical variational Bayesian source analysis technique revealed significantly greater event-related synchronization ERS and desynchronization ERD for correct over incorrect trials in the alpha, beta, and gamma frequency range prior to and after stimulus presentation. Together, these fMRI and MEG results are consistent with the hypothesis that articulatory processes serve to facilitate perceptual performance, while further dispelling concerns that activity found in ventral PMC/Broca's (mirror system) is merely a product of covert production of the perceived action. The finding of performance predictive activity prior to stimulus onset as well as activity related to task difficulty instead of information available in stimulation are consistent with constructivist and contrary to direct realist theories of perception.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The finding that premotor cortex PMC/Broca's areas are active not only during action production but also during visual and/or auditory observation of action ('Mirror Neurons' and 'mirror system') (Di Pellegrino et al., 1992; Gallese et al., 1996; Iacoboni and Dapretto, 2006; Kohler et al., 2002; Nishitani et al., 2005; Rizzolatti and Craighero, 2004; Rizzolatti and Arbib, 1998) has led to considerable conjecture regarding the neurophysiological mechanisms underlying a variety of abilities ranging from action perception to social cognition (Dapretto et al., 2006; Iacoboni, 2005; Nishitani et al., 2005). Two potential means by which the PMC/Broca's instantiates action recognition have been proposed and 'mirror' direct realist (Gibson, 1979) and constructivist (Helmholtz, 1867) theories of perception. The direct realist approach maintains that pickup of properties in sensory stimulation that have parity to self-generated actions affords perceptual recognition. It has been stated that neurons in the PMC / Broca's implement a non-inferential mechanism of action recognition

based on neural identity' (Iacoboni, 2005) or stated in other words action recognition emerges as a result of parity of the pattern of neural response for action observation and action generation. The constructivist approach maintains that analysis of ongoing sensory stimulation is constrained by a process of internal simulation that predicts sensory outcomes of action generation. Neurons in the PMC have been conjectured to instantiate forward internal models responsible for predicting sensory outcomes of action generation constraining perceptual recognition (Callan et al., 2004a; Wilson and Iacoboni, 2006; Iacoboni, 2008).

Direct realist and constructivist theories of speech perception are relevant in discerning the underlying processes involved with action recognition within PMC/Broca's. The use of articulatory goals to mediate perception was proposed long ago by the motor theory of speech perception (Liberman et al., 1967) and the theory of analysis by synthesis (Stevens and Halle, 1967). The main impetus for espousing the necessity to utilize articulatory constraints for speech perception was the apparent lack of invariant cues in the acoustic signal to specify our phenomenal experience; i.e. the d in 'dude' and in 'deep' sound alike but have different acoustic characteristics (Lotto et al., 2009; Galantucci et al., 2006; Liberman and Mattingly, 1985). The direct realist theory (Fowler, 1986) asserts that, although there are no acoustic/sensory

* Corresponding author. ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan. Fax: +81 774 95 1259.
E-mail address: dcallan@atr.jp (D. Callan).

features that invariantly specify the units of speech, there are invariant properties in sensory stimulation that unambiguously specify the articulatory gestures (instantiated in PMC/Broca's; my assertion) that are responsible for producing the units of speech in a direct manner without requiring the use of inferential processes. According to this approach what we perceive is not auditory/sensory in nature but directly relates to the articulatory gesture. On the other hand constructivist theories of speech perception maintain that speech perception is mediated by constraints of a model (inferential processes; internal simulation) of the articulatory–auditory (sensory) mapping of the speech production system. Examples of constructivist theories of speech perception include the motor theory of speech perception (Lieberman et al., 1967; Lieberman and Mattingly, 1985), analysis by synthesis (Stevens and Halle, 1967), and more recently internal models (Callan et al., 2003b, 2004a; Poeppel et al., 2008; Skipper et al., 2007b; Wilson and Iacoboni, 2006; Iacoboni, 2008). Internal models are mechanisms that simulate (predictively code) the input/output characteristics, or their inverses, of the system at hand (Kawato, 1999), in this case speech articulation and its sensory consequences. According to internal model based constructivist theories the PMC is involved with prediction of the sensory consequences of articulatory gestures that are used to constrain and facilitate speech perception. Although there are a considerable number of brain regions likely to be involved with speech perception this study focuses on the PMC and adjacent regions such as the opercular part of Broca's area because of the controversial involvement of the 'mirror system' in action recognition (Lotto et al., 2009; Hickok, 2008).

A considerable number of brain imaging studies have investigated the involvement of speech motor regions during auditory as well as visual speech perception (Callan et al., 2000, 2003a,b, 2004a,b, 2006a; Mottonen et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Nishitani and Hari, 2002; Pulvermuller et al., 2006; Skipper et al., 2005, 2007a; Wang et al., 2003; Wilson et al., 2004, there are far too many to list all). Most of these studies have only demonstrated that PMC and neighboring frontal opercular part of Broca's area are active both during speech perception tasks as well as speech production tasks. Although these findings are important they do not demonstrate that PMC/Broca's activity is related to perceptual performance. Demonstrating a relation to performance is necessary if these regions are truly involved with action recognition (Hickok, 2008). Without evidence that PMC/Broca's is related to perceptual performance it is entirely possible that the results reported in many previous studies merely reflect covert production of the perceived speech stimulus that is the product of perceptual processing and have no causal relationship with speech recognition. It should be noted that there are some studies that do show a relationship between PMC activity and perceptual performance. Greater activity is present in PMC/Broca's for perception of second language over native speech contrasts that increases with learning (Callan et al., 2006b, 2003a, 2004a; Wang et al., 2003). Furthermore, PMC/Broca's activity is correlated with subject mean behavioral identification performance on a second language phoneme identification task (Callan et al., 2004b). At first glance it would appear that the results of Binder et al. (2004) in which it was shown that Broca's activity increases with a reduction in performance is contrary to the results of the studies cited above. However, in the Binder et al. (2004) study stimuli were presented with different signal-to-noise ratios and performance decreased as the amount of auditory speech information was degraded. Consistent with the studies cited above, one could also interpret the results of greater activity in Broca's area (Binder et al., 2004) as reflecting facilitative processes used for speech perception under conditions when less auditory information is present.

Additionally, studies using transcranial magnetic stimulation TMS to speech motor areas (PMC and M1) have shown an effect on speech perception performance (Meister et al., 2007; D'Ausilio et al., 2009; Sato et al., 2009). However the results vary somewhat between studies.

Meister et al., 2007 find a decrease in phonetic discrimination performance with stimulation to PMv. D'Ausilio et al. (2009) find facilitation of performance with stimulation to M1. Sato et al. (2009) find a decrease in phonetic discrimination performance only when segmentation is required but not otherwise. There are considerable differences in the type of TMS used, the site stimulated, and the type of stimuli that may all contribute to differences in their results. One limitation of TMS methodology is that the potential spreading of activation induced by artificial stimulation of a particular brain region may result in distal changes in brain regions involved with perceptual performance that may not be influenced by the region undergoing TMS under normal circumstances. For example, stimulation of speech motor areas may activate forward models that are used for prediction of sensory consequences of the articulatory movement for speech production. These forward models originating in the PMC may have connections to auditory superior temporal areas but may not be used at all for speech perception under normal conditions but spreading activation caused by TMS unnaturally causes modulation of auditory speech maps resulting in facilitation or degradation in perceptual performance. It should be noted, however, that given the technique used in D'Ausilio et al.'s (2009) study it is unlikely that stimulation of speech motor areas modulate activity in the auditory cortex. Additionally Sato et al. (2009) did not show any effect of rTMS on phoneme identification and syllable discrimination tasks under normal listening conditions suggesting that stimulation may not result in spreading activation to auditory speech regions. To better corroborate the role of the PMC in perceptual processing it is necessary to demonstrate differential responses to correct and incorrect performance at an event level. Without evidence that PMC/Broca's processes correct and incorrect trials differently it is entirely possible that the results reported in many previous brain imaging studies merely reflect covert production of the perceived speech stimulus that is the product of perceptual processing. Although one cannot completely discard the possibility of a covert repetition strategy that is selective for correct responses we attempt to control for this possibility in this experiment in many ways such as comparing correct responses on a consonant identification task (difficult) versus correct responses on a vowel identification task (relatively easy) using identical syllables as stimuli. If greater activity for correct over incorrect responses is a product of covert production of the less ambiguous percept then one would expect the vowel task to also result in greater activity over the correct consonant task. Another way we address this issue is by looking at the timing of activity. At a minimum it would take at least 200 ms to listen to the speech stimuli and then be able to covertly repeat what was heard. Therefore if differences in activity are seen in PMC before or just after stimulus onset one can strongly rule out the difference in correct and incorrect trials as a product of covert production.

Here, we used functional magnetic resonance imaging fMRI and magnetoencephalography MEG to determine whether the same region of PMC/Broca's that is active during speech production is important for speech perception performance. This was accomplished by comparing brain activity for correct relative to incorrect trials on an auditory phonetic identification task in the presence of white noise. It is hypothesized that brain regions (PMC/Broca's) instantiating articulatory control signals (Callan et al., 2000; Pulvermuller et al., 2006; Wilson et al., 2004) will show greater activity for correct over incorrect presentations, reflecting its role in facilitation of perceptual processing. While fMRI allows for spatial localization of differential activity between correct and incorrect trials on the brain, MEG additionally allows for the time course of brain activity prior to and after stimulus presentation. Magnetoencephalography MEG, electroencephalography EEG, and electrocorticography ECoG studies have identified both event-related synchronization ERS as well as desynchronization ERD of activity in different brain areas than PMC/Broca's prior to stimulus onset to be predictive of performance (Ergenoglu et al., 2004; Gonzalez-Andino et al., 2005; Hanslmayr et al., 2007;

Linkenkaer-Hansen et al., 2004; Thut et al., 2006; van Dijk et al., 2008; Yamagishi et al., 2008). These results suggest that localized modulation of endogenous ongoing patterns of brain activity may be important for the effectiveness in which stimuli are processed. While the intent of this study is to investigate neural processes underlying the perceptual system one has to address the potential confound that cognitive decision processes are responsible for underlying brain activity. In an attempt to account for cognitive decision processes, analyses correlating brain activity with button response time (thought to be a correlate of decision processes (Binder et al., 2004)) were utilized to ensure that there was no overlap in brain activity related to perceptual performance.

We conducted three experiments (see **Materials and methods** for details) to test our hypothesis that PMC/Broca's is involved with processes related to perceptual performance: The first experiment is an fMRI study designed to localize the region of the PMC/Broca's that is active for both speech production of phonemes and speech perception of the same phonemes, specifying a potential mirror system site. These results are used for region of interest analyses in **Experiments 2 and 3**. The second experiment is an event-related fMRI study focusing on distinguishing correct from incorrect trials on a task involving two-alternative forced choice phonetic identification in the presence of white noise (see **Materials and methods** for details of stimulus presentation, the same procedure is used for **Experiments 2 and 3**). Activity found in the PMC/Broca's in **Experiment 1** common for both speech production and perception was used for an ROI analysis specifying the location of a potential mirror system site involved with perceptual performance. The third experiment uses the same phonetic identification task as the second experiment during MEG recording to identify the temporal oscillatory characteristics of activity in PMC/Broca's differentiating between correct and incorrect trials. Source localization is conducted using a hierarchical variational Bayesian technique (VBMEG; Sato et al., 2004; Yoshioka et al., 2008) that utilizes fMRI activity from experiment one as with prior information. Using this technique, current is extracted from PMC/Broca's known to be a potential mirror system site from which single trial time–frequency analyses are conducted.

The direct realist and constructivist theories of speech perception make different predictions regarding the outcome of the results of the various experiments. The direct realist theory supports the hypothesis that the PMC mediates action recognition by means of properties in sensory stimulation activating the same patterns of neural activity that are present during action production. The degree of information in stimulation that corresponds to the action should relate to the degree of neural activation in PMC. Additionally the activation in PMC is 'directly' related to properties in stimulation and thus is not involving predictive processing. Contrary to these hypotheses the constructivist theory of speech perception (in particular internal model based approaches) asserts that the PMC is involved with predictive processing (i.e. forward and inverse models) and the degree of activity may be dependent on the task demands rather than the amount of information available. The constructivist theory put forward here asserts that activity prior to stimulus onset may be predictive of performance. Additionally the constructivist theory asserts that task difficulty will reflect greater activity in PMC. Both of these predictions are contrary to the direct realist approach.

Materials and methods

Experiment 1: localization of brain regions involved with both production and perception

Subjects

Sixteen 21- to 42-year-old (14 males, 2 females) right-handed subjects participated in this study. All subjects gave written informed

consent for experimental procedures approved by the ATR Human Subject Review Committee.

Stimuli

The auditory stimuli consisted of the following synthesized male speech sounds /ba/, /bo/, /da/, and /do/ that were band passed filtered from 300 to 3400 Hz. The Festival speech synthesis system was used to construct the stimuli (Centre for Speech Technology Research, The University of Edinburgh). Each of the stimuli was 120 ms 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 44,100 Hz (Matlab was used for processing of the stimuli).

Procedure

The localization experiment consisted of four conditions: (1) overt speech production; (2) active perceptual identification of phonemes (/b/ versus /d/) in noise; (3) passive perception of phonemes; and (4) rest. The overt speech production condition consisted of articulation of one of the syllables (/ba/, /bo/, /da/, and /do/) presented in the center of the computer screen for 800 ms. To avoid large susceptibility artifact subjects were asked to articulate the speech sounds without voicing (overt articulation, but avoiding vibration of vocal folds). It was our opinion that the movement of the diaphragm and vocal folds during voicing would produce excessive head and body movement that may result in greater susceptibility artifact. Prior to the experiment all subjects demonstrated the ability to do this for all the speech stimuli without difficulty. An additional reason to exclude voicing was in order to avoid activation of auditory processing regions merely by acoustic feedback of one's own voice. The active perceptual identification of phonemes in noise condition consisted of two-alternative forced choice identification via button press by the left thumb of whether the syllable presented aurally begins with a /b/ or /d/. Subjects were informed to respond quickly and accurately. Button press order was counterbalanced. The speech syllables were embedded in 500 ms of white noise beginning at 200 ms. 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 passive perception of phonemes condition consisted of listening to the speech sound presented (/ba/, /bo/, /da/, and /do/) without making any overt response. The passive stimuli were presented at the same RMS energy as the combined speech and noise stimuli in the active speech condition. If an overt response was made in the passive condition the trial was excluded from the analysis.

The stimuli for each condition were presented in blocks of four trials. A color-coded fixation cross was presented during each block. The active conditions (speech production and active perceptual identification) were coded by a green cross, whereas, the passive conditions (passive speech perception and rest) were coded by a white cross. The first stimulus was presented 1000–1200 ms (randomly jittered) after the start of each block. Subsequent trials were presented 1800–2200 ms after the start of the previous stimulus. A single block was 10 s. There were 48 blocks total (12 repetitions for the 4 conditions) within a single run. Four runs were conducted. Subjects remained in the scanner between runs (resting for approximately 1 min) and were instructed not to move their head. The condition order was counterbalanced across runs and subjects. Subjects performed a practice run prior to fMRI scanning to familiarize them with the experiment.

fMRI data collection and preprocessing

The auditory stimuli were presented via MR-compatible Etymotic Research ER 30 Earphones (the delay in sound presentation resulting from the presentation computer sound board and the length of the air

tube were used to determine actual onset time). Visual stimuli were presented by projection via mirrors to a screen behind the head coil that could be viewed by the subject by a mirror. Stimulus presentation was controlled using Neurobehavioral System's Presentation software. The Siemens Trio 3 T was used for brain imaging at the ATR Brain Activity Imaging Center. Structural T1 images of the entire head were collected for each subject using a $1 \times 1 \times 1$ voxel resolution. Functional T2* weighted images were acquired using a gradient echo-planar imaging sequence (echo time 30 ms; repetition time 2000 ms; and flip angle 80°). A total of 20 interleaved axial slices were acquired with a $4 \times 4 \times 4$ mm voxel resolution covering the cortex and cerebellum. A total of 245 scans were taken for a single run. T2 structural images were also collected using the same horizontal slices as the functional image with a $1 \times 1 \times 4$ mm resolution. The experiment began from the start of the sixth scan. The first five scans were discarded. Images were preprocessed using programs within SPM2 (Wellcome Department of Cognitive Neurology, UCL). Differences in acquisition time between slices were accounted for and EPI images were unwrapped and realigned. The T2 image was coregistered to the mean EPI image, the EPI images were then spatially normalized to MNI space using a template T2 image ($3 \times 3 \times 3$ mm voxels) and the coregistered T2 image as the source. The images were smoothed using an $8 \times 8 \times 8$ mm FWHM Gaussian kernel. For use as constraints in the MEG analyses, the event-related analysis for each subject was also carried out without spatial normalization by coregistering the EPI images to each subject's T1 anatomical image and then smoothing using an $8 \times 8 \times 8$ mm FWHM Gaussian kernel. Regional brain activity for the various conditions was assessed using a general linear model employing an event-related design in which a modeled haemodynamic response function HRF was convolved with the event predictors. High pass filtering (cutoff period 128 s) was carried out to reduce the effects of extraneous variables (scanner drift, low frequency noise, etc.). Auto-regression was used to correct for serial correlations. Contrasts between the various conditions (speech production, active speech perception, passive speech perception, and rest) were conducted for each subject. Then random-effects analyses were conducted on the contrasts of interest involving the speech production, active speech perception, passive speech perception, and rest conditions. The statistics used in this experiment reflect a voxel level FDR threshold of $p < 0.05$ to correct for multiple comparisons over the entire brain. The cluster size specified by the voxel extent parameter in SPM was determined such that it had an uncorrected threshold of $p < 0.05$. The use of an extent threshold was to avoid the scattered activation of very small clusters across the brain. All coordinates reported in this study are given in Montreal Neurological Institute MNI units.

In fMRI experiments extensive head movement can produce excessive artifacts. This is of particular concern in speech production experiments in which the subject is moving their jaw, tongue, lips, etc.... In order to restrict head movement during the experiment the subject's head was securely fixed in place by using a strap that went across the forehead. The extent of scan-to-scan head movement in the x - y - z plane can be assessed using the realignment parameters given by SPM. The mean scan-to-scan movement for the trials used in the experiment was quite low (mean = 0.0755 mm, SD = .0567 mm). Although the speech production condition was significantly greater ($p < 0.05$) than the two perception conditions for 11 of the 13 subjects (Production: mean 0.0845 mm, SD = 0.0639 mm; Active perception: mean 0.0788 mm, SD = 0.0621 mm; Passive perception: 0.063 mm, SD = 0.0440; with the maximum being 0.25 mm across all subjects and conditions), it is unlikely that this small difference would have a significant effect on the brain imaging analysis given that the voxel size was 3 mm in the smallest dimension. An additional problem to be concerned with, especially in experiments using overt speech production is magnetic susceptibility artifacts (Yetkin et al., 1995). An event-related design (Palmer et al. 2001) and the use of unwarping during the realignment

step of preprocessing (SPM2) were utilized to minimize such magnetic susceptibility artifacts.

To investigate neural processes within particular theoretically relevant brain regions for Experiments 2 and 3 in which correct versus incorrect responses were evaluated, regions of interest ROIs were defined based on the results of Experiment 1. The four ROIs include the ventral inferior premotor cortex PMvi ($-51, 3, 18$), the ventral superior premotor cortex PMvs ($-54, -3, 39$), superior temporal gyrus STG ($-60, -24, 9$), and primary motor and somatosensory cortex M1/S1 (MNI 33, $-21, 54$). The PMvi and PMvs are brain regions that have been shown to be active in studies investigating speech production as well as speech perception (Callan et al., 2000; Wilson et al., 2004) demonstrating 'mirror system' properties. The STG is considered to be involved with auditory processing of speech and the M1/S1 region corresponds to the motor and sensory area representing the finger button press (Lotze et al., 2000). The STG and M1/S1 ROIs were utilized to ensure that activity localized to the PMvi and PMvs ROIs were not just a reflection of activity in these other regions arising from the auditory and/or motor responses. The PMvi and PMvs ROIs were determined by a conjunction analysis of activity common to the speech production relative to rest contrast, the active speech perception relative to rest contrast, and the passive speech perception relative to rest contrast. The conjunction analysis was carried out by finding the intersection of significantly active voxels between these contrasts of interest (determined at $pFDR < 0.05$) (based on method given in Nichols et al., 2005). A series of contrasts was conducted to determine task related activity in the STG and M1/S1 that could be used to define the ROIs. The STG ROI was determined by voxels present for both the active speech perception and passive speech perception relative to rest contrasts ($pFDR < 0.05$) that were also significantly greater for active speech perception over passive speech perception ($pFDR < 0.05$) and active speech perception over speech production ($pFDR < 0.05$). The M1/S1 ROI was determined by voxels present for the active speech relative to rest contrast ($pFDR < 0.001$) that were also significantly greater for active speech perception relative to speech production ($pFDR < 0.05$) and active speech perception relative to passive speech perception ($pFDR < 0.05$). It was necessary to use a strict statistical threshold ($pFDR < 0.001$) for the active speech relative to rest contrast in order to reduce the cluster size in M1/S1 to be focused on the area representing the finger for the button press.

Results and discussion

Behavioral performance

Performance on the perceptual identification task (/b/ versus /d/) in the presence of white noise was evaluated by percent correct trials. Experiment 1; Mean d prime = 1.60; SE = 0.20; Mean percent correct = 75.2%; SE = 2.4%, (significantly greater than chance; $t = 10.7$; $p < 0.0001$; $df = 15$).

Brain imaging

The brain imaging results for the speech production relative to rest contrast, the active speech perception relative to rest contrast, and the passive speech perception relative to rest contrast are given in Figs. 1A–C (Production: $T = 2.55$, $pFDR < 0.05$, spatial extent threshold = 80 voxels; Active: $T = 2.64$, $pFDR < 0.05$, spatial extent threshold = 59 voxels; Passive: $T = 3.27$, $pFDR < 0.05$, spatial extent threshold = 35 voxels; spatial extent is selected based on uncorrected cluster level $p < 0.05$). Fig. 1D and Table 1 give the results of a conjunction analysis showing activity common to the speech production contrast, the active speech perception contrast, and the passive speech perception contrast based on the method given in Nichols et al. (2005) using the intersection of significant voxels

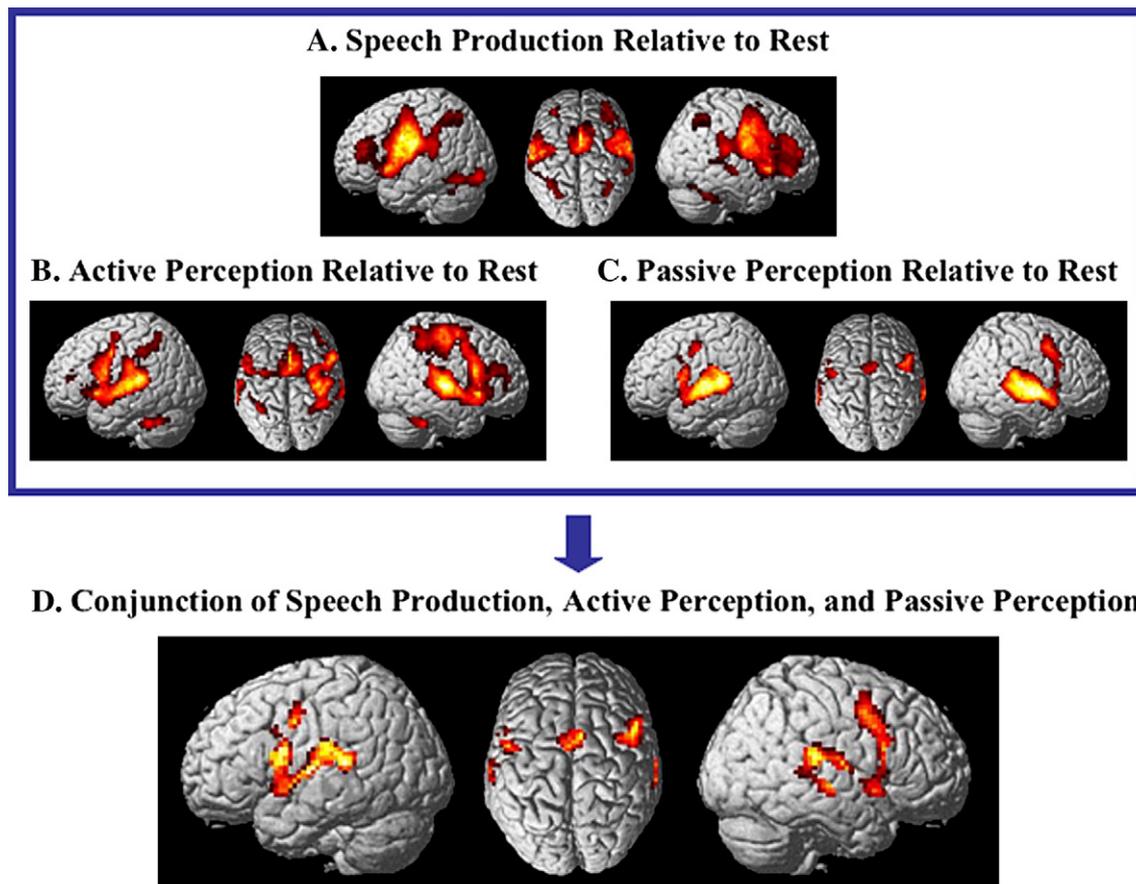


Fig. 1. Result of random-effects fMRI analysis for Experiment 1. A. Speech production condition relative to rest condition. B. Active speech perception condition relative to rest condition. C. Passive speech perception condition relative to rest condition. D. Conjunction analysis showing activity that is common to all three experimental conditions. All results are shown using a threshold of $pFDR < 0.05$.

($pFDR < 0.05$) of the results given in Figs. 1A–C. Brain regions showing common activity for production and perception tasks include the inferior ventral premotor cortex PMvi including Broca's area, the superior ventral PMvs, the anterior insula, the supplementary motor area SMA, the superior temporal gyrus STG and transverse temporal gyrus, and the inferior parietal lobule IPL (Fig. 1D, Table 1).

With regards to evaluating the influence that speech motor areas have on perception, we focused on brain regions denoted by activity in the PMvi and PMvs, as they are sites referred to in the literature as

showing 'mirror system' properties that are found to be active during a variety of speech perception tasks (Callan et al., 2000, 2003a,b, 2004a,b, 2006a; Mottonen et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Nishitani and Hari, 2002; Pulvermuller et al., 2006; Skipper et al., 2005, 2007a; Wang et al., 2003; Wilson et al., 2004). Regions of interest ROIs, used in further fMRI and MEG analysis (Experiments 2 and 3), were constructed based on the results of Experiment 1 (see Materials and methods for details). These ROIs include PMvi and PMvs (hypothesized mirror system) (Figs. 2A–B), left superior temporal gyrus STG (auditory processing) (Fig. 2C), as well as primary motor and somatosensory cortex M1/S1 of the finger corresponding to the button press (Fig. 2D). It should be noted that the series of contrasts used to find the clusters of voxels composing the ROI also resulted in activity elsewhere in the brain as can clearly be seen by comparing Figs. 1D and 2A–B.

Experiment 2: fMRI analysis of correct relative to incorrect performance

Subjects

Thirteen 25- to 42-year-old (12 males, 1 females) right-handed subjects participated in this study. Three of these subjects also participated in Experiments 1 and 3. All subjects gave written informed consent for experimental procedures approved by the ATR Human Subject Review Committee.

Stimuli and procedure

The same stimuli and procedures were used for Experiment 2 and Experiment 3. Both experiments (2 and 3) consisted of three conditions: (1) active perceptual identification of consonant phonemes (/b/ versus /d/) in noise; (2) active perceptual identification of vowel phonemes (/a/ versus /o/) in noise; and (3) passive

Table 1

Experiment 1: MNI coordinates of peak activity of clusters.

Conjunction of speech production, active perception, and passive perception contrasts ($pFDR < 0.05$)	
Brain region	MNI coordinate
Anterior insula	45,9,−6
PMvi BA6 and 44 (including Broca's area)	−51,3,18 48,12,18
PMvs BA 6	−54,−3,39 51,9,36
SMA BA 6	−6,9,48 6,6,54
STG BA22/Transverse temporal gyrus BA42	−63,−27,12 −57,−33,9 −48,−12,6 −51,3,−6 63,−15,9 65,−27,15 53,−21,−3
Inferior parietal lobule post central gyrus BA40/43	−60,−27,21 63,−30,21

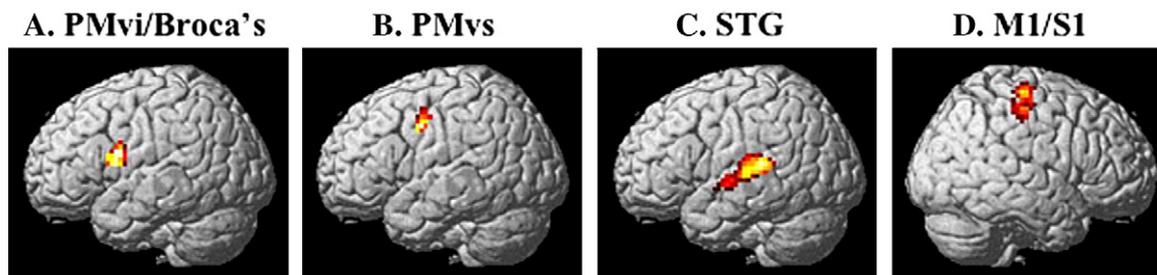


Fig. 2. Regions of interest ROIs where MEG current was localized and submitted to time–frequency analysis. A. Ventral premotor cortex including the opercular part of Broca's area PMC/Broca's PMvi. B. Ventral superior premotor cortex PMvs. C. Superior temporal gyrus STG. D. Primary motor and somatosensory cortex M1/S1.

listening to noise with no speech presented. The active perceptual identification conditions consisted of a two-alternative forced choice perceptual identification task (/b/ versus /d/) or (/a/ versus /o/) in the presence of white noise by means of a button press by the left thumb. Subjects were informed to respond quickly and accurately. Button press order was counterbalanced. The white noise was presented 1000–2500 ms (randomly jittered) prior to the onset of the same phonetic speech sounds used in [Experiment 1](#) (/ba/, bo/, /da/, /do/). The white noise remained present for 1500 ms after initial onset of the speech sound (see [Fig. 3](#) for depiction of stimulus presentation procedure). 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. In the passive listening to noise without speech condition subjects were instructed to listen to the white noise while fixating on the cross presented on the center of the screen and make no button response. The duration of the white noise was matched to that of the active perceptual condition (2500–4000 ms).

The stimuli for each condition were presented in blocks of six trials. The instructions as to the task condition were given visually at the start of each block for 1200 ms. The block duration was fixed at 38 s. The time between trials within a block was 2500–3000 ms (randomly jittered). The active perceptual identification condition was presented after each subsequent block of either /vowel/ or /passive/ trials. There were 8 blocks in each run (4 active perception /b/-/d/, 2 passive listening to noise, and 2 active perception /a/-/o/) and a total of 8 runs for the entire experiment. Subjects remained in the scanner between runs (resting for approximately 1 min) and were instructed not to move their head. The order of the blocks was counterbalanced across runs.

fMRI data collection and preprocessing

The auditory stimuli were presented via MR-compatible Hitachi ceramic transducer headphones (the delay in sound presentation resulting from the presentation computers sound board was used to determine actual onset time). Visual stimuli were presented by projection via mirrors to a screen behind the head coil that could be viewed by the subject by a mirror. Stimulus presentation was controlled using Neurobehavioral System's Presentation software. 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 55 ms; repetition time 2000 ms; flip angle 90°). A total of 20 sequential axial slices were acquired with a 3.5 × 3.5 × 6 mm voxel resolution (one mm gap) covering the cortex and cerebellum. A total of 152 scans were taken for a single run. 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 MNI space using a template EPI image (3 × 3 × 3 mm voxels), and were smoothed using a 7 × 7 × 12 mm FWHM Gaussian kernel. Regional brain activity for the various conditions was assessed using a general linear model employing an event-related design in which a modeled haemodynamic response function HRF was convolved with the event predictors. High pass filtering (cutoff period 128 s) was carried out to reduce the effects of extraneous variables (scanner drift, low frequency noise, etc.). Auto-regression was used to correct for serial correlations. Contrasts between the following four conditions were conducted for each subject: correct consonant perceptual identification (Correct); incorrect consonant perceptual identification (Incorrect); correct vowel perceptual identification (Vowel); and passive listening to noise (pasnoise). For each subject and run the number of correct trials (consonant task) and vowel trials (only correct responses) was matched to the number of incorrect trials (consonant task). A minimum of three incorrect trials per run was necessary for inclusion. Selection of trials was randomly made for each subject based on matching for contrast type (/ba-da/ or /bo-do/) across the incorrect, correct, and vowel conditions and duration of noise before the stimulus was presented. Only 3 out of the 13 subjects had significant differences in button response time (the time from the onset of the stimulus till the button press) between correct and incorrect conditions. However, 10 subjects out of the 13 had significant differences in button response time between the correct and vowel conditions as well as the incorrect and vowel conditions (see [Results and discussion](#)). To reduce possible contamination of activity related to the button response their onset times were convolved with a haemodynamic response and used as a regressor of non-interest in the analysis. Random-effects analyses were conducted on the contrasts of interest (Correct, Incorrect, Vowel, and pasnoise). The statistics used in the experiments reflect a voxel level FDR threshold of $p < 0.05$ to correct for multiple comparisons over the entire brain. In

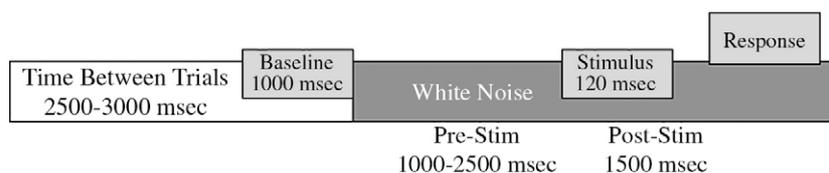


Fig. 3. Stimulus presentation procedure for Experiments 2 and 3. The speech stimulus (/ba/, /da/, /bo/, or /do/) was aurally presented randomly between 1000 and 2500 ms after the onset of white noise. The white noise remained on for 1500 ms after the onset of the stimulus. The total duration of the white noise therefore ranged from 2500 ms to 4000 ms. The stimuli were 120 ms in duration. The task for the subject was to determine if the stimulus begins with a /b/ or a /d/ and to respond with the left hand by button press. The time between trials ranged from 2500 to 3000 ms. The 1000 ms before the white noise was presented was used as baseline for MEG analysis.

the case where the FDR threshold was not significant an uncorrected threshold of $p < 0.005$ was used and additional theoretically motivated region of interest analyses were conducted using a small volume correction for multiple comparisons with a FDR threshold of $p < 0.05$. The cluster size specified by the voxel extent parameter in SPM was determined such that it had an uncorrected threshold of $p < 0.05$. The use of an extent threshold was to avoid the scattered activation of very small clusters across the brain.

The extent of scan-to-scan head movement in the x - y - z plane between the three experimental conditions (correct, incorrect, and vowel) was assessed. There was no significant difference between the various conditions for all subjects (Correct: mean = 0.1122, SD = .0791; Incorrect: mean = 0.1150, SD = 0.0671; Vowel: mean = 0.1180, SD = 0.0805; with the maximum being 0.25 mm across all subjects and conditions) with the exception of just a single subject showing a significant difference for the correct relative to vowel contrast. Magnetic susceptibility artifacts were minimized by the use of an event-related design (Palmer et al. 2001) and the use of unwarping during the realignment step of preprocessing (SPM2).

Results and discussion

Behavioral performance

Performance on the perceptual identification task (consonant: /b/ versus /d/) and (vowel: /a/ versus /o/) in the presence of white noise was evaluated by percent correct trials. Consonant identification task: Mean d prime = 1.76; SE = 0.19; Mean percent correct = 79.1%; SE = 2.8%, (significantly greater than chance; $p < 0.05$; $df = 12$). Vowel identification task: Mean d prime = 4.35; SE = 0.40; Mean percent correct = 96.2%; SE = 1.3%, (significantly greater than chance; $p < 0.05$; $df = 12$). Behavioral performance for the consonant identification task was significantly lower than that of the vowel identification task (paired $T = 5.4$; $p < 0.05$; $df = 12$). There was no significant correlation between behavioral performance (Mean percent correct 79.1%, SE 2.8%) for each subject and the difference in correct minus incorrect button response times on the consonant identification task (Correct: mean 1.164 s, SE 0.071 s; Incorrect: mean 1.097 s, SE 0.087 s) ($r = 0.21$, $p > 0.1$). Neither was there a significant correlation between the difference in behavioral performance of the consonant minus vowel tasks (Mean percent correct: Consonant = 79.1%, SE = 2.8%; Vowel = 96.2%, SE = 1.3%) and the difference in button response times for consonant (Mean response time = 1.11 s; SE = 0.079 s) minus vowel (Mean response time = 0.870 s; SE = 0.046 s) tasks ($r = -0.33$, $p > 0.1$).

Brain imaging

The random-effects brain imaging results for the correct relative to pasnoise contrast, the incorrect relative to pasnoise contrast, and the vowel relative to pasnoise contrast are given in Figs. 4A–C and Table 2 (Correct: $T = 3.06$, $pFDR < 0.05$, spatial extent threshold = 35 voxels; Incorrect: $T = 23.61$, $pFDR < 0.05$, spatial extent threshold = 25 voxels; Vowel: $T = 3.20$, $pFDR < 0.05$, spatial extent threshold = 28 voxels; spatial extent is selected based on uncorrected cluster level $p < 0.05$). All three contrasts show very similar patterns of activity including the PMvi, PMvs, anterior insula, SMA, M1/S1, STG, IPL, basal ganglia, brain stem, and thalamus (Table 2). The random-effects brain imaging results for the contrasts of correct relative to incorrect and correct relative to vowel are shown in Figs. 5A–B (Correct–Incorrect: $T = 3.05$, $p < 0.005$ uncorrected, spatial extent threshold = 37 voxels; Correct–Vowel: $T = 3.05$, $p < 0.005$ uncorrected, spatial extent threshold = 33 voxels). To avoid erroneous conclusions based on differences in subthreshold activity, only voxels that were significantly active for the correct relative to the control pasnoise condition (Fig. 4A) were included in the results (Figs. 5A–B). Additional region of interest analyses were carried out using small volume correction for

multiple comparisons within the sites defined in Experiment 1 (combined PMvi and PMvs Figs. 2A–B; STG Fig. 2C; M1/S1 Fig. 2D). Significant differential activity between correct and incorrect conditions was present in the combined PMvi and PMvs ROI defined by activity present for both speech production and perception tasks in Experiment 1 ($T = 4.15$, $pFDR < 0.05$, MNI coordinate $-54, 3, 27$; $T = 3.42$, $pFDR < 0.05$, MNI coordinate $-54, 0, 36$, Table 3). The small volume correction region of interest analysis for the correct relative to the vowel contrast also revealed significant differential activity in the PMvi/PMvs region ($T = -45, 6, 15$, $pFDR < 0.05$, MNI coordinate $-45, 6, 15$, Table 3). No significant differential activity for small volume correction analyses was present in the STG or M1/S1 regions of interest (Figs. 2C–D) for the correct relative to incorrect, or correct relative to vowel contrasts (Figs. 5A–B). At first it may seem that our results are at odds with those Binder et al. (2004) in which greater activity was found to correlate with phonetic identification performance in STG. However, in their study the signal-to-noise ratio of the stimuli were varied and highly correlated with behavioral performance (better performance for stimuli with higher signal-to-noise ratio). Therefore one could also interpret their result of greater activity in STG to reflect processes involved with the greater degree of complex acoustic feature information that is present in stimuli with higher signal-to-noise ratios.

To ensure that the results of the fMRI analyses reported above (Figs. 4–5) were not reflecting cognitive decision processes a correlation analysis between button response time (thought to reflect cognitive decision making processes (Binder et al., 2004)) and brain activity was performed. There were no statistically significant ($pFDR > 0.05$, $p > 0.005$ uncorrected, spatial extent threshold $p > 0.05$ uncorrected) positive or negative correlations between button response time and brain activity for correct relative to pasnoise (correct button response time regressor), incorrect relative to pasnoise (incorrect button response time regressor), vowel relative to pasnoise (vowel button response time regressor), correct relative to incorrect contrasts (correct minus incorrect button response time regressor) and correct relative to vowel contrasts (correct minus vowel button response time regressor).

The ventral PMC region (PMvi/PMvs) found in our study to differentiate between correct and incorrect trials (Fig. 5A) and correct and vowel identification (Fig. 5B) have been shown in previous studies to be active during a variety of tasks involving speech perception (Basirat et al., 2008; Callan et al., 2003a, 2000, 2004a; Skipper et al., 2005; Wang et al., 2003) and speech production (Bohland and Guenther, 2006; Callan et al., 2000; Wilson et al., 2004). Although we find peak activity in the PMvs in the same region ($-46, -6, -45$) as found in Wilson et al. (2004) ($-51, -11, 46$) it is unclear why their study did not also report activity in PMvi including the opercular part of Broca's area as we found in this study. The inferior ventral PMC and the opercular part of Broca's area is more in line with area F5 in the non-human primate studies demonstrated to contain 'Mirror Neurons' considered to be the analogue of Broca's area (Rizzolatti and Arbib, 1998; Rizzolatti and Craighero, 2004). It is also adjacent to the motor region representing the speech articulators (Pulvermuller et al., 2006). The Wilson et al. (2004) study does report speech production peak activity in the PMvi at ($-56, -4, 22$) which is near the peak activity found in our study ($-48, 6, 14$) from the conjunction analysis (Fig. 1D; Table 1). It is entirely possible that the Wilson et al. (2004) study will reveal significant activity in the PMvi if a region of interest analysis is conducted centered on the peak activity found in our study. Several other studies have also found activity present in the PMvi during tasks involving non-native phonetic perception (Callan et al., 2006b, 2003a, 2004a; Wang et al., 2003) and visual speech gesture perception (Callan et al., 2003b; Paulesu et al., 2003; Calvert and Campbell, 2003; Bernstein et al., 2002; Nishitani and Hari, 2002; Olson et al., 2002; Campbell et al., 2001; Skipper et al., 2005; only to name a few). However, there are many studies investigating speech perception that do not find activity in

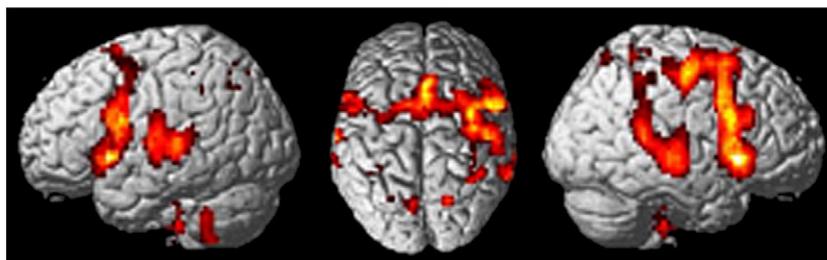
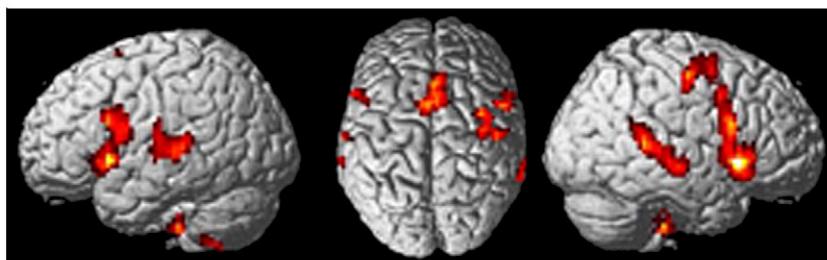
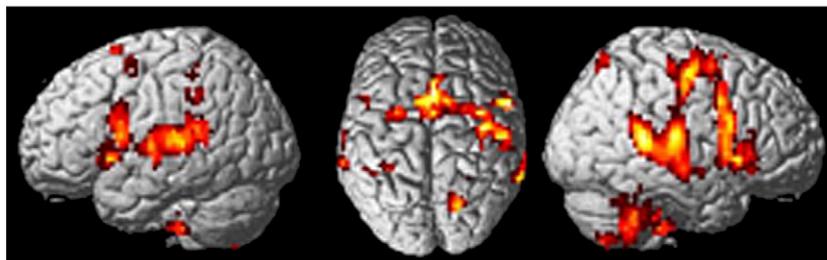
A. Correct Relative to PasNoise**B. Incorrect Relative to PasNoise****C. Vowel Relative to PasNoise**

Fig. 4. Result of random-effects fMRI analysis for Experiment 2. A. Correct relative to the pasnoise condition. B. Incorrect relative to the pasnoise condition. C. Vowel relative to the Pas Noise condition. All results are shown using a threshold of $pFDR < 0.05$.

this region (Callan et al., 2006a, 2004b; MacSweeney et al., 2001; Calvert et al., 1997, 1999, 2000; Wilson et al., 2004; Wilson and Iacoboni, 2006; only to name a few). Although the tasks and conditions vary widely throughout the many studies one general trend that may be responsible for the finding of PMvi involvement in some studies and not in others is the use of active tasks (consistent with findings of Gold et al., 2005, that this region is involved with controlled phonetic processing). It may be the case that the PMvi including Broca's is only active when predictive coding is utilized. Our study included both active phonetic identification and passive listening to phonemes presented as monosyllables whereas for example the Wilson et al. (2004) study only involved passive listening to monosyllables. Although our passive conditions are quite similar the inclusion of an active phonetic identification task in our study may have engaged processes related to predictive coding to a greater extent than in Wilson et al. (2004). While it is also possible that brain activity found in the anterior insula, SMA, and inferior parietal lobule in the passive task (as well as in the conjunction analysis of all three tasks in Experiment 1) may be in part a product of covert production we do not believe the differences in correct and incorrect trials and the differences in correct versus vowel trials in Experiment 2 are merely a result of covert production (see general discussion).

Experiment 3: MEG analysis of correct relative to incorrect performance

Subjects and stimuli

Subjects used in Experiment 3 were the same as those used in Experiment 1. The order in which the subject participated in the fMRI or MEG experiment was randomly determined.

The stimuli and experimental procedure was the same as that of Experiment 2 (see Fig. 3) with the exception that in the active speech condition a lower signal-to-noise ratio was used because of the additional noise during fMRI scanning that was not present during MEG recording. 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 2 dB signal-to-noise ratio.

MEG data collection and preprocessing

The auditory stimuli were presented via MR-compatible Etymotic Research ER 30 Earphones (the delay in sound presentation resulting from the presentation computers sound board and the length of the air tube were used to determine actual onset time). Visual stimuli were presented by projection to a screen in front of the subject. Stimulus presentation was controlled using Neurobehavioral System's Presentation software. The Yokogawa 208 channel MEG supine position system at the ATR Brain Activity Imaging Center was used to collect the data. A sampling rate of 1000 Hz was used with input gain of $\times 5$ and an output gain of $\times 100$ using a high pass filter of 0.3 Hz and a low pass filter of 200 Hz. To restrict head movement the subjects were restrained using a strap across the forehead.

For each trial the MEG data was segmented 1000 ms prior to the onset of the noise (baseline) and from 1000 ms pre and post onset of the speech stimulus. For the passive noise condition stimulus onset was determined to be 1500 ms prior to the end of the presentation of the white noise. EEGLAB (Delorme and Makeig, 2004) was used to preprocess the MEG data. The baseline was subtracted from the trials and independent component analysis was conducted using principle

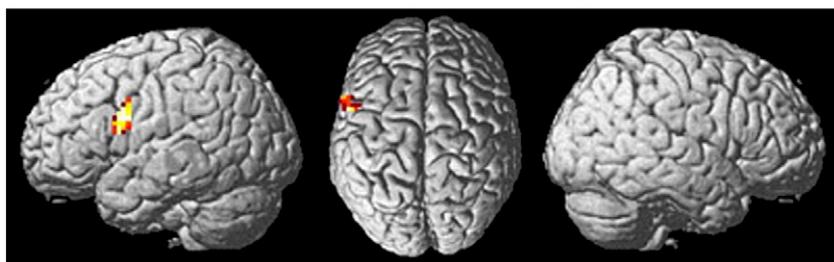
Table 2
Experiment 2: MNI coordinates of peak activity of clusters ($pFDR < 0.05$).

Brain region	Correct versus pasnoise	Incorrect versus pasnoise	Vowel versus pasnoise
	MNI coordinate	MNI coordinate	MNI coordinate
Basal ganglia	–21,9,–3 –12,3,3 21,3,–3	–21,9,–6 21,9,3	–12,0,–3 21,6,3 27,–3,0
Brain stem	0,–33,–42	–3,–33,–42	0,–33,–45
Thalamus	9,–15,3	6,–12,0	15,–18,3
Anterior insula	–36,12,3 –33,18,–3 36,18,–6	–36,12,3 39,15,–6	–36,12,3 39,24,6
PMvi BA6 and 44 (including Broca's area)	–48,9,15 42,12,21	–48,3,18 51,9,18	–48,6,18 45,9,18
PMvs BA 6	–54,3,42 54,6,39	–49,9,33 54,6,42	–48,6,33 54,6,36
Dorsal PMC	–30,–3,51	–6,6,57	–27,0,57
SMA BA 6/Pre-SMA	6,27,48	6,12,54 6,24,48	–6,9,63 9,9,54
M1/S1	36,–12,54	48,–18,48	39,–12,54
STG BA22, transverse	–63,18,12	–63,–18,12	–57,–30,9
temporal gyrus BA42,	–45,–30,3	–63,–24,6	63,–42,6
MTG BA21	63,–42,12 63,–33,3 57,–24,–9	63,–39,6 57,–21,–9 66,–48,18	57,–18,–9 60,–21,9
Inferior parietal lobule BA40/43	–63,–18,24 –36,–39,45	–63,18,21	52,–22,22
Supramarginal gyrus	–36,–48,33		–42,–45,36 60,48,22
Superior parietal lobule	–27,–63,39 36,–51,60 18,–66,54	–28,–60,42	
Parietal lobe precuneus	–12,–72,42		–6,–69,48 15,–57,51
Cerebellum	–15,–57,–21 –3,–60,–27 –27,–60,–36 –39,–48,–51		–6,63,–21 –15,–54,–21 –27,–60,–33 –45,–45,–45 24,–66,–54 39,–54,–45 51,–51,–36

component reduction to 52 components. In total 8 runs using the same procedure as Experiment 2 were recorded. Subjects remained in the scanner between runs (resting for approximately 1 min) and were instructed not to move their head. During this time measures of head position were taken. Because there was very little head movement within the MEG all 8 runs were concatenated for analysis (the mean maximum head movement deviation throughout the 8 runs across the 16 subjects was 1.8 mm with SD = 0.7 mm and a maximum value of 2.9 mm). Artifact rejection was carried out over the independent component activation waveforms. The rejection of entire components was determined by first finding components with over 20 trials with outlier values (using a SD threshold of 10) or abnormal linear trends (using a slope window width of 1000 points, a maxslope of 50, and a minR of .3) and then inspecting the components topographic scalp map as well as trial activity across time and the power spectrum for characteristics of normal artifacts such as eye movement, eye blinks, muscle activity, etc. Components deemed to be artifacts were removed (the median number of components removed from each subject was 8). Individual trials with outlier values (SD threshold of 20) were also removed from the data. Further analysis was conducted over the same number of trials selected from the correct active perception condition, the incorrect active perception condition, and the passive noise condition. Trials were selected to balance for each subject the number of items for each contrast type /ba-da/ and /bo-do/, the number of stimuli per run, the onset time, and the button response time ($p > 0.1$ two-tailed for all subjects). Once these constraints were taken into account trial selection was randomized.

The position of the head within the MEG was determined by the use of five coils attached to the subject's head (one behind each ear, and three across the forehead). The positions of the markers for the coils on the face were measured by the Polhemus FastSCAN Cobra system. This system obtains a 3D laser scan of the face as well as the coordinate location for the five markers. In house Matlab software was used to register the coordinate space of the 3D face image to the subject's anatomical T1 MRI structural image. Given the position of the 5 coils in reference to the MEG sensors the position of these sensors can be registered in the coordinate space of the subject's T1 MRI structural image. Brain Voyager software was used for

A. Correct Relative to Incorrect Perception



B. Correct Relative to Vowel Perception

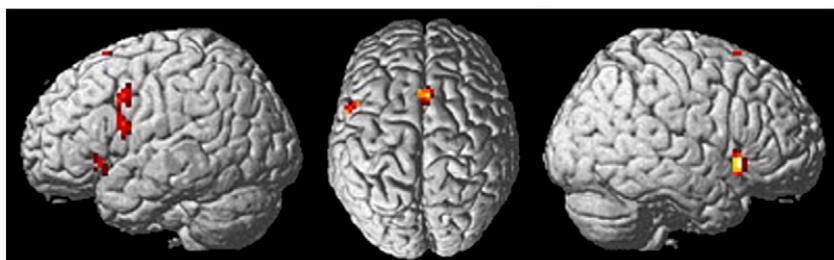


Fig. 5. fMRI results for correct perception relative to incorrect perception conditions. A. Significant differential brain activity rendered on the surface of the brain for correct relative to incorrect trials is present in PMC/Broca's ($p < 0.005$ uncorrected over entire brain; Region of interest analysis over PMC reveals activity is significant at $pFDR < 0.05$). B. Significant differential brain activity rendered on the surface of the brain for correct relative to vowel trials ($p < 0.005$ uncorrected over entire brain; Region of interest analysis over PMC reveals activity is significant at $pFDR < 0.05$).

Table 3
Experiment 2: MNI coordinates of peak activity of clusters.

Brain region	Correct versus incorrect	Correct versus vowel
	MNI coordinate	MNI coordinate
Basal ganglia	12,6,3	
Anterior insula		–30,24,0
		42,15,–3
PMvi BA6 and 44 (including Broca's area)	–54,3,27 ^a	–45,6,15 ^a
PMvs BA 6	–54,0,36 ^a	–57,3,39
		–51,9,36
SMA BA 6		6,18,60

^a Denotes activity significant at $pFDR < 0.05$ using a small volume correction for multiple comparisons within the region of interest of combining PMvi and PMvs given in Figs. 4A–B.

segmentation of the cortex in order to obtain vertex points for the MEG leadfield model within the same coordinate space as the MRI anatomical image for which the MEG sensor positions have been coregistered. The leadfield model was estimated using the Sarvas equation (Sarvas, 1987). Source localization was conducted using a variational Bayesian technique that utilizes fMRI prior information as constraints (VBMEG). See Sato et al. (2004), and Yoshioka et al. (2008), for a detailed description of VBMEG as well as its tested accuracy of source localization.

The fMRI data of Experiment 1 was used to define constraints for VBMEG analysis as well as to specify ROIs (Figs. 1 and 2). The T map for each subject's specific fMRI event-related analysis of the active perception condition relative to rest (Experiment 1: Threshold at $p < 0.005$ uncorrected, spatial extent = 25 voxels) was used as constraints for VBMEG analysis (VBMEG parameters: variance magnification factor = 100, confidence parameter = 500, these are the same values as used in Yoshioka et al., 2008; windows for analysis were from –2000 to –1501, –1750 to –1251, –1500 to –1001, –1000 to –501, –750 to –251, –500 to –1, –250 to 249, 0 to 499, 250 to 749, and 500 to 999; baseline period –2000 to –1001; and stimulus onset at 0). VBMEG analysis estimated current activity over the entire cortex.

The individual brain anatomy specific voxels (specifying the vertex points for single trial VBMEG analysis) for the regions of interest including PMvi, PMvs, STG, and M1/S1 (Given in Figs. 2A–D) were determined using the following method: The coordinates of the ROI from the random-effects analysis (Figs. 2A–D) were projected onto each subject's cortex using their individual specific normalization parameters given by SPM. The intersection of these coordinates with the fixed effect individual specific fMRI activity for the active perception condition relative to rest ($p < 0.005$) is then determined to define the vertex values of the ROI for each individual on their own brain anatomy. The PMvi and PMvs ROIs had the additional constraint of requiring fixed effect individual specific fMRI activity for the speech production condition relative to rest ($p < 0.005$).

Time–frequency analyses were carried out using event-related spectral perturbation ERS (EEGLAB, Delorme and Makeig, 2004). The single trial current for all three conditions (correct perception, incorrect perception, passive noise) from –1000 to +1000 ms for the vertex point in each ROI with the largest mean post onset value were entered into the time–frequency analysis (time–frequency analysis parameters: wavelet 0-padded DFTs, 3 cycles at lowest frequency (11.7 Hz) to 25.6 cycles at highest frequency (200 Hz); window size was 285 samples; time–frequency matrix 100×49 pixels denoting 11.7 to 200 Hz and –857 to +857 ms). The time range from –250 to 650 ms and the frequency range from 11 to 130 Hz (31×53 pixels) were extracted from the time–frequency analysis results for further statistical analysis. The mean value across time for the passive noise condition for each subject was used as a baseline and

statistically subtracted from the correct perception and incorrect perception conditions (Paired t -test, $N = 16$). Additionally, the correct perception and incorrect perception conditions were statistically compared (Paired t -test, $N = 16$).

Results and discussion

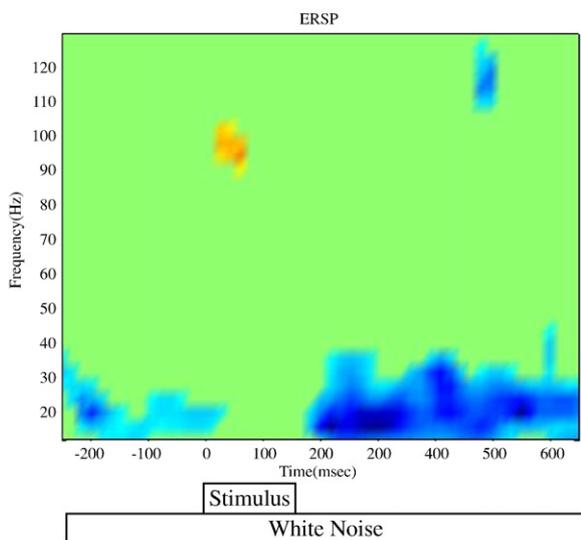
Behavioral performance

Performance on the perceptual identification task (consonant: /b/ versus /d/) and (vowel: /a/ versus /o/) in the presence of white noise was evaluated by percent correct trials. Consonant identification task: Mean d prime = 1.11; SE = 0.10; Mean percent correct = 70.1%; SE = 1.6%, (significantly greater than chance; $p < 0.05$; $df = 15$). Vowel identification task: Mean d prime = 3.54; SE = 0.30; Mean percent correct = 94.0%; SE = 1.1%, (significantly greater than chance; $p < 0.05$; $df = 15$). Behavioral performance for the consonant identification task was significantly lower than that of the vowel identification task (paired $T = 5.9$; $p < 0.05$; $df = 15$). Similar to Experiment 2, there was no significant correlation between behavioral performance (Mean percent correct 70.1%, SE 1.6%) for each subject and the difference in correct minus incorrect button response times on the consonant identification task (Correct: mean 0.941 s, SE 0.037 s; Incorrect: mean 0.978 s, SE 0.036 s) ($r = -0.003$, $p > 0.1$). Neither was there a significant correlation between the difference in behavioral performance of the consonant minus vowel tasks (Mean percent correct: Consonant = 70.1%, SE = 1.6%; Vowel = 94.0%, SE = 1.1%) and the difference in button response times for consonant (Mean response time = 0.83 s; SE = 0.033 s) minus vowel (Mean response time = 0.960 s; SE = 0.036 s) tasks ($r = -0.05$, $p > 0.1$).

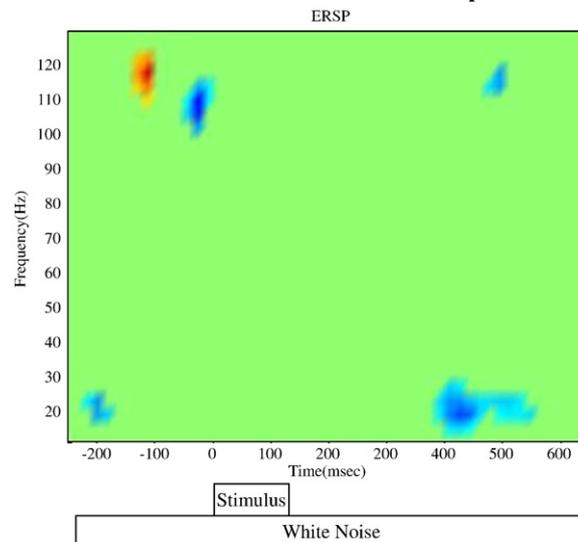
Brain imaging

The results of the time–frequency analysis (ERSP) of the MEG current localized to the PMvi (Fig. 2A) is given in Figs. 6A–C. The correct perception relative to the pasnoise contrast showed significant ($pFDR < 0.05$ two-tailed paired t -test; corrected across entire matrix 31×53) event-related desynchronization ERD peaks around 200 ms prior to stimulus onset as well as 200 ms after stimulus onset extending out to 650 ms in the alpha, beta, and low gamma range 11.7 to 35 Hz as well as a high frequency 115 Hz peak ERD around 500 ms (Fig. 6A). Additionally, a significant event-related synchronization ERS occurring around 50 ms post stimulus for the correct perception condition was found in the 100 Hz range (Fig. 6A). For the incorrect perception relative to the pasnoise contrast ($pFDR < 0.05$ two-tailed paired t -test; corrected across entire matrix 31×53) peak ERD are present at 350 and 570 ms after stimulus onset in the alpha, beta, and low gamma range 11.7 to 35 Hz (Fig. 6B). For the vowel perception relative to the pasnoise contrast ($pFDR < 0.05$ two-tailed paired t -test; corrected across entire matrix 31×53) event-related desynchronization ERD peaks at 150 ms after stimulus onset and at 250 ms after onset extending out to 600 ms in the alpha, beta, and low gamma range 11.7 to 35 Hz (Fig. 6C). Significant event-related synchronization ERS peaks occurred around 280 ms post stimulus around 105 Hz and at 600 ms around 60 Hz (Fig. 6C). To ensure that the results of the ERS analysis reported above (Figs. 6–7) were not reflecting cognitive decision processes a correlation analysis between button response time (thought to reflect cognitive decision making processes (Binder et al., 2004)) and brain activity was performed. There were no significant positive or negative correlations ($pFDR > 0.05$) between ERS and button response time for the correct relative to pasnoise (correct button response time regressor), incorrect relative to pasnoise (incorrect button response time regressor), vowel relative to pasnoise (vowel button response time regressor), as well as correct relative to incorrect contrasts (correct minus incorrect button

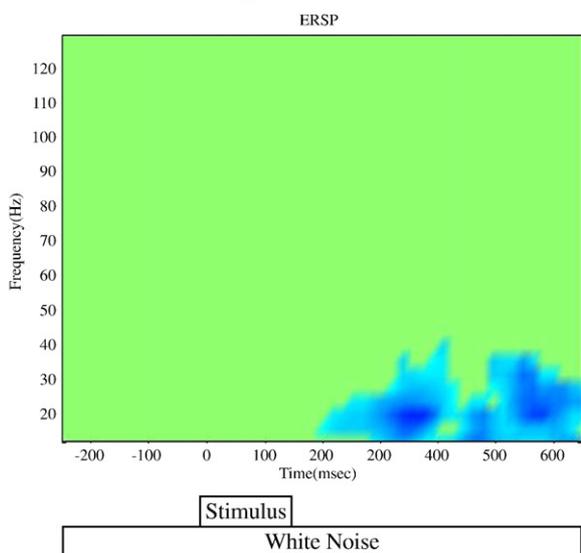
A. Correct Perception Relative to PasNoise



Correct Relative to Incorrect Perception



B. Incorrect Perception Relative to PasNoise



C. Vowel Perception Relative to PasNoise

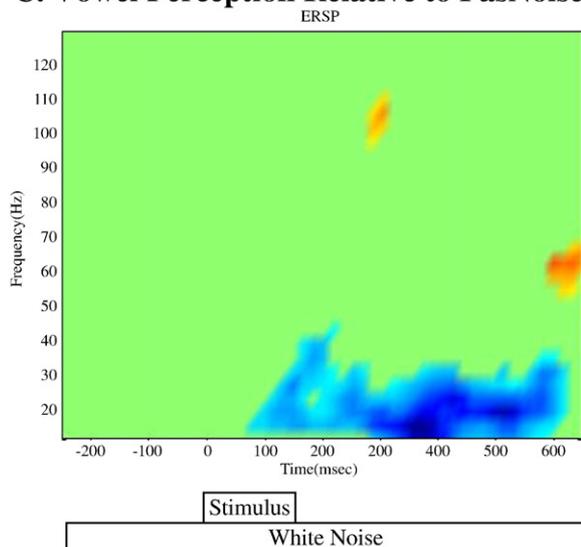


Fig. 7. Random-effects results of the MEG time–frequency event-related spectral perturbation ERSP analysis across single trials within PMvi for Correct relative to Incorrect trials. Dark colors on plot denote activity significant at $pFDR < 0.05$; light colors denote activity significant at $p < 0.05$ uncorrected.

response time regressor) in any of the regions of interest (PMvi, PMvs, STG, and M1/S1).

The contrast of primary interest, correct perception relative to incorrect perception, is given in Fig. 7. When correcting for multiple comparisons across the entire matrix ($pFDR < 0.05$ two-tailed) a significant ERS was found to occur in the 100 Hz range 100 ms prior to stimulus onset and a significant ERD was found immediately prior to stimulus onset in the 110 Hz range. Using the combined activity from the correct perception and incorrect perception contrasts (at $p < 0.0034$ uncorrected; value of uncorrected p corresponding to $pFDR$ for the correct relative to pasnoise contrast) as a ROI, a SVC analysis was conducted comparing the correct perception relative to the incorrect perception condition (Fig. 7; $pFDR < 0.05$ two-tailed; uncorrected $p < 0.0026$). The results of the SVC analysis show significantly greater ERD 200 ms prior to stimulus onset and 400 ms to 570 ms post stimulus in the high alpha, beta and low gamma range 14–32 Hz. Additional ERD occurs at 500 ms post stimulus onset in the 115 Hz range (Fig. 7). The time–frequency analyses carried out between the correct perception and incorrect perception conditions for the PMvs (Fig. 2B), STG (Fig. 2C) and M1/S1 (Fig. 2D) did not reveal any significant activity using the same analysis procedure as conducted for the PMvi. The ERSP analysis of the correct relative to vowel contrast did not reveal any significant activity ($pFDR > .05$) for PMvi, PMvs, STG, or M1/S1.

While fMRI provides excellent spatial resolution its temporal resolution is somewhat limited. EEG and MEG allows for oscillatory patterns of activity to be investigated both before and after stimulus presentation. Several EEG, and MEG studies have identified that prestimulus activity is related to attention and behavioral perceptual performance. A decrease in alpha power (ERD) prior to stimulus onset is predictive of better visual perception performance (Hanslmayr et al., 2007; van Dijk et al., 2008; Yamagishi et al., 2008). It has been proposed that alpha activity may reflect an inhibitory filter mechanism mediated by top-down attentional control (Klimesch et al.,

Fig. 6. Random-effect results of the MEG time–frequency event-related spectral perturbation ERSP analysis across single trials within PMvi: A. Correct relative to pasnoise condition. B. Incorrect relative to pasnoise condition. C. Vowel relative to pasnoise condition. Red denotes event-related synchronization; Blue denotes event-related desynchronization. Dark colors on plot denote activity significant at $pFDR < 0.05$; light colors denote activity significant at $p < 0.05$ uncorrected.

2007). A release from inhibition, signified by an ERD, allows for improved processing by the spreading of activity that serves to ready the system for coding of incoming information (Klimesch et al., 2007; Pfurtscheller and Lopes da Silva, 1999; Yamagishi et al., 2005, 2008). In our study, a significant ERD was seen prestimulus as well as after the offset of stimulus presentation for both correct and incorrect trials (Figs. 6A–B). The alpha and beta ERD 200 ms prior to stimulus onset as well as 400–550 ms after stimulus onset were found to be significantly larger for correct over incorrect trials (Fig. 7). Our results are consistent with the hypothesis that greater disinhibition of alpha activity in PMC/Broca's results in activation of control processes involved with phoneme articulation used to facilitate speech perception.

EEG studies have also shown that high frequency activity (>40 Hz) can predict performance (Gonzalez-Andino et al., 2005; Hanslmayr et al., 2007). In our study we find a significant 100 Hz ERS approximately 50 ms after stimulus onset for correct trials (relative to the passive listening to noise condition) (Fig. 6A), whereas for incorrect trials no significant ERS was found. Correct trials additionally showed a significant 115 Hz ERD (relative to the passive listening to noise condition) approximately 500 ms after stimulus onset (Fig. 6A). For correct over incorrect trials we find significant differential ERS and ERD prior to stimulus presentation as well as an ERD 600 ms after stimulus presentation in PMC/Broca's (Fig. 7). Our results are consistent with previous studies showing the involvement of high frequency gamma activity in the inferior frontal gyrus (overlapping with the PMC/Broca's found in our study) during speech perception and production of vowels (Callan et al., 2000) as well as for the verbal transformation speech effect for two patients implanted with electrodes (Basirat et al., 2008). Gamma activity is thought to mediate both feed-forward bottom up processing of information important for binding of perceptual features (Hanslmayr et al., 2007) as well as, especially for very high frequency activity (>80 Hz), top-down attentional control related to stimulus discrimination (Lenz et al., 2008) and selective attention (Ray et al., 2008) that is performance related (Gonzalez-Andino et al., 2005). In agreement with these studies, the greater ERD and ERS for correct over incorrect trials found in our study, may reflect attentional modulation related to selection and inhibition of articulatory neural processes that place constraints on perception and facilitate performance.

The lack of a significant difference between correct and incorrect trials in STG and M1/S1 would tend to shed doubt on the hypothesis that differential activity in the PMC/Broca's was just an artifact of current arising from these regions. Additionally, high frequency activity associated with motor responses only occurs over contralateral sensorimotor cortex (Crone et al., 1998). In our study, subjects responded with their left thumb therefore, one would not expect the left PMC to be active during such a movement. Additional support that our MEG analysis reflects differences in activity in the PMC and not from some artifact or activity arising from elsewhere is that the results of the fMRI study (Experiment 2: Fig. 5A) only revealed activity in the PMC and nowhere else in the brain to differentiate correct from incorrect trials.

Discussion

The results of the three experiments conducted in this study strongly suggest that the ventral PMC including the opercular part of Broca's area identified as having mirror system properties is an influential part of the perceptual system that is involved in facilitating performance. Both the fMRI (Experiment 2; Figs. 4–5) and MEG results (Experiment 3; Figs. 6–7) indicate significant differential activity for correct over incorrect trials within the ventral PMC/Broca's area. This region was found by conjunction analysis to be present during speech production, active speech perception in which a button response was required for phonetic identification, and during

passive speech perception of the same phonemes (Experiment 1; Fig. 3). The presence of neural activity during action observation and action execution are hallmarks of the mirror system (Iacoboni, 2005; Iacoboni and Dapretto, 2006; Rizzolatti and Craighero, 2004). These results are highly relevant in that they dispel concerns that the activity found in brain regions involved with motor planning and execution during perception are merely a result of covert production of the perceived speech stimulus arising from a perceptual system upon which no influence is exerted. Additionally the results have ramifications with respect to predictions made by direct realist and constructivist theories of speech perception.

Several results in this study lend support to constructivist theories of speech perception and are somewhat contrary to direct realist theories of speech perception. The constructivist theory asserts that the 'mirror system' will be engaged to a greater extent in predictive coding the more difficult the task. In this experiment, the white noise degraded consonant information much more than vowel information (This is reflected in the overall behavioral results during fMRI, Consonant task: 79.1%; Vowel task: 96.2%). Therefore, the constructivist theory predicts that the correct condition should show greater 'mirror system' activity than the vowel condition. In contrast, the direct realist theory asserts that the 'mirror system' will be engaged in relation to the amount of information in sensory stimulation that affords the speech gesture. Therefore, accordingly the vowel condition should show greater 'mirror system' activity than the correct condition. The results of Experiment 2 show that PMC/Broca's is more active for the correct relative to incorrect contrast as well as for the correct relative to vowel contrast (Figs. 5A–B). These results support the constructivist theory of speech perception suggesting that the 'mirror system' is engaged differentially based on the extent that articulatory–auditory (sensory) information can be used as a predictive constraint to facilitate perception.

The constructivist theory asserts that prediction based on articulatory constraints are used to facilitate speech perception (top-down or recurrent processing), therefore one may expect the system to be active prior to stimulus onset. The direct realist theory asserts that the configuration of the properties in sensory stimulation specify the afforded percept (bottom up processing), therefore one would only expect activity concurrent with stimulus onset. Consistent with the prediction of constructivist theory the results of the time-frequency analysis within PMvi indicate significant differential activity between correct and incorrect trials prior to stimulus onset (Fig. 7) suggesting that the mechanism of the mirror system may involve predictive coding consistent with utilization of an internal model (Callan et al., 2003a,b, 2004a; Iacoboni and Wilson, 2006a,b; Iacoboni, 2008). It has been proposed that the motor system (more specifically the mirror system) may be utilized for perceptual anticipation and prediction to facilitate performance (Blackemore and Frith, 2005; Chaminade et al., 2001; Kilner et al., 2007; Lamm et al., 2007; Schubotz, 2007). It should be noted that although there is considerable evidence that PMC/Broca's is involved with articulatory planning the utilization of an internal model to carry out this constructive process is just one of many possible mechanisms that may be utilized.

We maintain that the primary finding in our study of greater PMC/Broca's activity for correct over incorrect trials reflects greater involvement of the 'mirror system' in facilitating action recognition. However, alternatively, it may be possible that the incorrect trials reflect a failure of the auditory perceptual system such that the subject will perform the required button press without performing any covert production of a percept as is done for correct trials. One could conjecture that correct trials tax more working memory and subvocal rehearsal strategies resulting in covert production which is known to activate similar speech motor regions as was found in our study (Rogalsky et al., 2008). There are several reasons why we do not believe this to be the case: 1. No significant differential activity

between correct and incorrect trials was found in auditory perceptual regions for the fMRI (Fig. 5A) or the MEG experiments. If the response in Broca's/PMC was merely covert production of a correct percept of the auditory system one would expect brain regions involved with auditory aspects of speech processing to show differential activity between correct and incorrect trials. 2. The MEG results show activity in PMvi differentiating correct versus incorrect responses before as well as early after stimulus onset (Fig. 7). These results suggest that the 'mirror system' may be involved with predictive coding such as that of an internal model. 3. There is no reason to assume that phenomenological ambiguous speech percepts denoting incorrect trials would be covertly produced less than trials with more phenomenological apparent speech percepts. In fact research on perception of non-native phonetic contrasts suggests greater activity in PMC/Broca's as well as auditory speech processing regions (Callan et al., 2003a, 2004a; Wang et al., 2003). 4. The contrast of the correct versus the vowel condition (also reflecting correct responses) showing significant differential activity in the PMC/Broca's (Fig. 5B) suggests that the results are not just merely a reflection of covert production of phenomenological apparent speech percepts.

The results of this experiment strongly suggest that there is a motor component to speech perception. This is exemplified by greater activity for correct over incorrect responses found in both fMRI and MEG studies (Figs. 5A and 7). One could interpret these results as support for the position that recruitment of predictive articulatory-based processes is fundamental for perceptual success. Correct responses associated with greater activity utilize these processes whereas incorrect responses where perception fails do not utilize these processes. However, activity in PMvi does not appear to be modulated just by intelligibility but also based on task demands. This is shown by greater activity for the consonant correct relative to vowel correct contrast (Fig. 5B). These results would seem to be consistent with the hypothesis that articulatory-based processes are only used when there is insufficient bottom up information available for purely perceptual processes to succeed (contrary to stronger motor theory positions stating that it is necessary for perception). Consistent with this hypothesis Binder et al. (2004) showed that by decreasing the signal-to-noise ratio of the stimuli that the relative contributions of the auditory area reflecting complex acoustic speech feature processing decreased (as did intelligibility) and conversely that activity in Broca's area increased perhaps reflecting greater processing demands of predictive articulatory-based processes. However, one cannot, completely discount the possibility that articulatory processes are necessary for speech perception based on our results alone as activity was present in speech motor areas (PMvi) for all conditions including passive and active consonant as well as vowel conditions for all our experiments (Figs. 1, 4–7). Additionally while our results suggest that processing in PMvi is task dependant the activity is also important for perceptual performance and is not merely a result of differential activation of general cognitive mechanisms involved with subvocal rehearsal (see arguments above). Neither can the activity in PMvi be explained by greater decision processes independent of performance as a correlation between brain activity and button response time (considered to reflect decision processes (Binder et al., 2004)) was not found to be statistically significant.

The task dependent nature of activity in the PMvi suggests that the articulatory-based processes that facilitate perception are used to a greater extent when the auditory information is degraded. One may use this task dependency to claim that articulatory-based processes are not fundamental to speech perception but only help to facilitate auditory-based speech perception under certain conditions (Lotto et al., 2009). Supporters of this position would claim that the true fundamental processes involved with speech perception are discerned by listening to clear speech presented without noise (so called normal speech). However, in fact 'normal' speech perception that we experience in

real-world conditions almost always occurs in the presence of background noise. The neural processes whether they be domain specific or general that evolved to accommodate speech perception were shaped based in an environment full of noise rather than in silence. Therefore, consistent with the results of our experiment in which speech intelligibility was only found to be predicted by activity in PMC and not in auditory areas (Figs. 5A and 7), one could conjecture that articulatory-based processes are fundamental to speech perception under ordinary listening conditions in which noise is present. Further research needs to be conducted to determine the extent to which PMC mediate and/or facilitate speech perception under silent background conditions.

Given the task dependent nature of the processing of auditory and speech motor regions to speech intelligibility the question arises as to what mediates the contribution of the various systems. It is likely that both bottom up stimulus driven processes as well as top-down attention processes are utilized. Consistent with the theory of neuronal group selection (Edelman, 1987, 1989; Edelman and Tononi, 2000), it is conjectured that perceptual categorization is made possible by global mappings composed of degenerate neural mapping systems (both auditory- and motor-based, that are localized in different brain regions) as well as limbic and value based neural systems. The degenerate auditory-based and motor-based neural systems process the information available in stimulation differently. It is the properties of the stimulus itself that serve to select in part which degenerate neural processes are predominantly used for perception. It is further hypothesized that task dependent attention can modulate selection and inhibition of the various degenerate neural systems. Our finding of a significantly greater alpha ERD for correct over incorrect trials before stimulus onset is consistent with an attentionally modulated filter mechanism (Klimesch et al., 2007) that modulates (via a release from inhibition) selection and inhibition of articulatory neural processes that place constraints on perception and facilitate performance.

The results of the three experiments of this study strongly support the hypothesis that the brain regions involved with action production are an integral part of the perception system serving to facilitate performance. Our results have far reaching implications bringing for the first time performance related support to brain imaging studies proposing that action perception is mediated by brain regions involved with action production. Furthermore, our results represent a key piece of evidence validating previous research espousing the use of brain regions responsive to action production and perception (mirror system) as instantiating neurophysiological mechanisms underlying a variety of abilities ranging from the embodiment of language (Glenberg and Kaschak, 2002; Glenberg et al., 2008; Rizzolatti and Arbib, 1998), and cognition (Garbarini and Adenzato, 2004) to explanation of disorders such as autism (Dapretto et al., 2006). The results of our experiments are consistent with constructivist type processes (internal model) and not a result of direct perception (direct realism). The finding that ventral PMC/Broca's activity differentiates correct from incorrect trials prior to stimulus onset is consistent with the view that the mirror system instantiates internal models involved in anticipatory prediction and selection in order to support performance (Kawato, 1999; Oztop et al., 2005; Wolpert et al., 2003). The ability to predict future performance based on prestimulus activity has implications for brain machine interfaces in which manipulative neuroscience could be used to enhance behavioral and learning performance.

Acknowledgments

We thank Ray Kent and Nicolas Schweighofer for their comments in preparing this manuscript as well as Taku Yoshioka and Ryosuke Hayashi for assistance with MEG analysis. We also thank fMRI and MEG technicians Yasuhiro Shimada, Ichiro Fujimoto and Hironori Nishimoto at the Brain Activity Imaging center at ATR as well as Yuka Furukawa for assisting in running the experiments. This research was

supported by a contract with the National Institute of Information and Communications Technology entitled, 'Multimodal integration for brain imaging measurements' and by KAKENHI, Grant-in-Aid for Scientific Research(C) (21500321).

References

- Basirat, A., Sato, M., Schwartz, J., Kahane, P., Lachaux, J., 2008. Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *NeuroImage* 42 (1), 404–413.
- Bernstein, L., Auer, E., Moore, J., Ponton, C., Don, M., Singh, M., 2002. Visual speech perception without primary auditory cortex activation. *NeuroReport* 13, 311–315.
- Binder, J., Liebenthal, E., Possing, E., Medler, D., Ward, D., 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301.
- Blackmore, S.J., Frith, C., 2005. The role of motor contagion in the prediction of action. *Neuropsychologia* 43, 260–267.
- Bohland, J., Guenther, F., 2006. An fMRI investigation of syllable sequence production. *NeuroImage* 32, 821–841.
- Callan, D., Callan, A., Honda, K., Masaki, S., 2000. Single-sweep EEG analysis of neural processes underlying perception and production of vowels. *Cogn. Brain Res.* 10, 173–176.
- Callan, D., Tajima, K., Callan, A., Kubo, R., Masaki, S., Akahane-Yamada, R., 2003a. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19, 113–124.
- Callan, D., Jones, J., Munhall, K., Callan, A., Kroos, C., Vatikiotis-Bateson, E., 2003b. Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* 14 (17), 2213–2218.
- Callan, D., Jones, J., Callan, A., Akahane-Yamada, R., 2004a. 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.
- Callan, D., Jones, J., Munhall, K., Kroos, C., Callan, A., Vatikiotis-Bateson, E., 2004b. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J. Cogn. Neurosci.* 16, 805–816.
- Callan, D., Tsytarev, V., Hanakawa, T., Callan, A., Katsuhara, M., Fukuyama, H., Turner, R., 2006a. Song and speech: brain regions involved with perception and covert production. *NeuroImage* 31, 1327–1342.
- Callan, A., Callan, D., Tajima, K., Akahane-Yamada, R., 2006b. Neural processes involved with perception of non-native durational contrasts. *NeuroReport* 17, 1353–1357.
- Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cogn. Neurosci.* 15, 57–70.
- Calvert, G.A., Bullmore, E., Brammer, M.J., Campbell, R., Williams, S., McGuire, P., Woodruff, P., Iversen, S., David, A., 1997. Activation of auditory cortex during silent lipreading. *Science* 276, 593–596.
- Calvert, G.A., Brammer, M., Bullmore, E., Campbell, R., Iversen, S.D., David, A., 1999. Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport* 10, 2619–2623.
- Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G.A., McGuire, P., Suckling, J., Brammer, M.J., David, A.S., 2001. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cogn. Brain Res.* 12, 245–264.
- Chaminade, T., Meary, D., Orliaguet, J., Decety, J., 2001. Is perceptual anticipation a motor simulation? A PET study. *NeuroReport* 17, 3669–3674.
- Crone, N., Miglioretti, D., Gordon, B., Lesser, R., 1998. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121, 2301–2315.
- Dapretto, M., Davies, M., Pfeifer, J., Scott, A., Sigmann, M., Bookheimer, S., Iacoboni, M., 2006. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat. Neurosci.* 9 (1), 28–30.
- D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, L., Beglioni, C., Fadiga, L., 2009. The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180.
- Edelman, G., 1987. *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books, New York.
- Edelman, G., 1989. *The Remembered Present: A Biological Theory of Consciousness*. Basic Books, New York.
- Edelman, G., Tononi, G., 2000. *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books, New York.
- Ergenoglu, T., Demirlp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., Uresin, Y., 2004. Alpha rhythm of the EEG modulates visual detection performance in humans. *Brain Res. Cogn. Brain Res.* 20, 376–383.
- Fowler, C., 1986. An event approach to the study of speech perception from a direct-realist perspective. *J. Phon.* 14, 3–28.
- Galantucci, B., Fowler, C., Turvey, M., 2006. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13 (3), 361–377.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119 (Pt2), 593–609.
- Garbarini, F., Adenzato, M., 2004. At the root of embodied cognition: cognitive science meets neurophysiology. *Brain Cogn.* 56 (1), 100–106.
- Gibson, J.J., 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.
- Glenberg, A., Kaschak, M., 2002. Grounding language in action. *Psychon. Bull. Rev.* 9 (3), 558–565.
- Glenberg, A., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., Buccino, G., 2008. Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61 (6), 905–919.
- Gold, B., Balota, D., Kirchoff, B., Buckner, R., 2005. Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from fMRI adaptation. *Cereb. Cortex* 15, 1438–1450.
- Gonzalez-Andino, S., Michel, C., Thut, G., Landis, T., Grave de Peralta, R., 2005. Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain. *Human Brain Mapp.* 24, 50–58.
- Hanslmayr, S., Aslan, A., Staudigl, T., Kilimesch, W., Herrmann, C., Bauml, K., 2007. Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage* 37, 1465–1473.
- Helmholtz, H., 1867. *Handbuch der physiologischen Optik*. L. Voss, Leipzig.
- Hickok, G., 2008. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21 (7), 1229–1243.
- Iacoboni, M., 2005. Understanding others: imitation, language, empathy. In: Hurley, S., Chater, N. (Eds.), *Perspectives on Imitation: From Cognitive Neuroscience to Social Science*, Vol. 1: Mechanisms of Imitation and Imitation in Animals. MIT Press, Cambridge, MA, pp. 77–99.
- Iacoboni, M., 2008. The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *J. Physiol. Paris* 102, 31–34.
- Iacoboni, M., Wilson, S., 2006a. Beyond a single area: motor control and language within a neural architecture encompassing Broca's area. *Cortex* 42 (4), 503–506.
- Iacoboni, M., Wilson, S., 2006b. Beyond a single area: motor control and language within a neural architecture encompassing Broca's area.
- Iacoboni, M., Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951.
- Kawato, M., 1999. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727.
- Kilner, J., Friston, K., Frith, C., 2007. The mirror–neuron system: a Bayesian perspective. *NeuroReport* 18, 619–623.
- Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88.
- Kohler, E., Keyser, C., Umiltà, M., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848.
- Lamm, C., Fischer, M.H., Decety, J., 2007. Predicting the actions of others taps into one's own somatosensory representations—a functional MRI study. *Neuropsychologia* 45, 2480–2491.
- Lenz, D., Jeschke, M., Schadow, J., Naue, N., Ohl, F., Herrmann, C., 2008. Human EEG very high frequency oscillations reflect the number of matches with a template in auditory short-term memory. *Brain Res.* 1220, 81–92.
- Lieberman, A., Mattingly, I., 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Lieberman, A., Cooper, F., Shankweiler, D., Studdert-Kennedy, M., 1967. Perception of speech code. *Psychol. Rev.* 74, 431–461.
- Linkenkaer-Hansen, K., Nikulin, V., Palva, S., Ilmoniemi, R., Palva, J., 2004. Prestimulus oscillations enhance psychophysical performance in humans. *J. Neurosci.* 24 (45), 10186–10190.
- Lotto, A., Hickok, G., Holt, L., 2009. Reflections on mirror neurons on speech perception. *Trends Cogn. Sci.* 13 (3), 110–114.
- Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., Grodd, W., 2000. fMRI evaluation of somatotopic representation in human primary motor cortex. *NeuroImage* 11, 473–481.
- MacSweeney, M., Campbell, R., Calvert, G., McGuire, P., Davie, A., Suckling, J., Andrew, C., Woll, B., Brammer, M.J., 2001. Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people. *Proc. R. Soc. Lond. Series B Biol. Sci.* 268, 451–457.
- Meister, I., Wilson, S., Deblieck, C., Wu, A., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696.
- Mottonen, R., Jarvelainen, J., Sams, M., Hari, R., 2004. Viewing speech modulates activity in the left S1 mouth cortex. *NeuroImage* 24, 731–737.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J., 2005. Valid conjunction inference with the minimum statistic. *NeuroImage* 25, 653–660.
- Nishitani, N., Hari, R., 2002. Viewing lip forms: cortical dynamics. *Neuron* 36, 1211–1220.
- Nishitani, N., Schurmann, M., Amunts, K., Hari, R., 2005. Broca's region: from action to language. *Physiology* 20, 60–69.
- Ojane, V., Mottonen, R., Pekkol, J., Jaaskelainen, I., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *NeuroImage* 25 (2), 333–338.
- Olson, I.R., Gatenby, J.G., Gore, J.C., 2002. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cogn. Brain Res.* 14, 129–138.
- Oztop, E., Wolpert, D.M., Kawato, M., 2005. Mental state inference using visual control parameters. *Cogn. Brain Res.* 22, 129–151.
- Palmer, E., Rosen, H., Ojemann, J., Buckner, R., Kelley, W., Petersen, S., 2001. *NeuroImage* 14 (1 Pt 1), 182–193.
- Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N.A., De Giovanni, U., Sensolo, S., Fazio, F., 2003. A functional-anatomical model for lip-reading. *J. Neurophysiol.* 90 (3), 2005–2013.

- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I., Mottonen, R., Tarkiainen, A., Sams, M., 2005. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *NeuroReport* 16 (2), 125–128.
- Pfurtscheller, K., Lopes Da Silva, F., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857.
- Poeppel, D., Idsardi, W.J., van Wassenhove, V., 2008. Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 12 (363(1493)), 1071–1086.
- Pulvermuller, F., Huss, M., Kheri, F., Moscoso, Del Prado, Martin, F., Hauk, O., Shtrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci.* 103, 7865–7870.
- Ray, S., Niebur, E., Hsiao, S., Sinai, A., Crone, N., 2008. High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention. *Clin. Neurophysiol.* 119, 116–133.
- Rizzolatti, G., Arbib, M., 1998. Language within our grasp. *Trends Neurosci.* 21 (5), 188–194.
- Rizzolatti, G., Craighero, L., 2004. The mirror–neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rogalsky, C., Matchin, W., Hickok, G., 2008. Broca's area, sentence comprehension, and working memory: an fMRI study. *Front. Hum. Neurosci.* 2 (14), 1–13.
- Sato, M., Yoshioka, T., Kajiwara, S., Toyama, K., Goda, N., Doya, K., Kawato, K., 2004. Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage* 23, 806–826.
- Sato, M., Tremblay, P., Gracco, V., 2009. A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111 (1), 1–7.
- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* 32, 11–22.
- Schubotz, R., 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218.
- Skipper, S., Nusbaum, H., Small, S., 2005. Listening to talking faces: motor cortical activation during speech perception. *NeuroImage* 25, 76–89.
- Skipper, J., Goldin-Meadow, S., Nusbaum, H., Small, S., 2007a. Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277.
- Skipper, J., van Wassenhove, V., Nusbaum, H., Small, S., 2007b. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399.
- Stevens, K.N., Halle, M., 1967. Remarks on analysis by synthesis and distinctive features. In: Walther-Dunn, W. (Ed.), *Models for the Perception of Speech and Visual Form*. MIT Press, Cambridge, MA.
- Thut, G., Nietzel, A., Brandt, S., Pascual-Leone, A., 2006. Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26 (37), 9494–9502.
- van Dijk, H., Schoffelen, J., Oostenveld, R., Jensen, O., 2008. Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823.
- Wang, Y., Sereno, J., Jongman, A., Hirsch, J., 2003. fMRI evidence for cortical modification during learning of mandarin lexical tone. *J. Cogn. Neurosci.* 15, 1019–1027.
- Wilson, S., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage* 33, 316–325.
- Wilson, S., Saygin, A., Sereno, M., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.
- Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Biol. Sci.* 358, 593–602.
- Yamagishi, N., Goda, N., Callan, D., Anderson, S., Kawato, K., 2005. Attentional shifts towards an expected visual target alter the level of alpha-band oscillatory activity in the human calcarine cortex. *Cogn. Brain Res.* 25, 799–809.
- Yamagishi, N., Callan, D., Anderson, S., Kawato, M., 2008. Attentional changes in pre-stimulus oscillatory activity within early visual cortex are predictive of human visual performance. *Brain Res.* 1197, 115–122.
- Yetkin, F., Hammeke, T., Swanson, S., Morris, G., Mueller, W., McAuliffe, T., Houghton, V., 1995. A comparison of functional MR activation patterns during silent and audible language tasks. *Am. J. Neuroradiol.* 16 (5), 1087–1092.
- Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., Sato, M., 2008. Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals. *NeuroImage* 42, 1397–1413.