HHS Public Access

Author manuscript

J Cogn Neurosci. Author manuscript; available in PMC 2017 January 18.

Published in final edited form as:

J Cogn Neurosci. 2014 March ; 26(3): 606–620. doi:10.1162/jocn_a_00515.

Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect and fMRI

William Matchin, **Kier Groulx**, and **Gregory Hickok**

Department of Cognitive Sciences, Center for Cognitive Neuroscience, University of California, Irvine

Abstract

Visual speech influences the perception of heard speech. A classic example of this is the McGurk effect, whereby an auditory /pa/ overlaid onto a visual /ka/ induces the fusion percept of /ta/. Recent behavioral and neuroimaging research has highlighted the importance of both articulatory representations and motor speech regions of the brain, particularly Broca's area, in audiovisual (AV) speech integration. Alternatively, AV speech integration may be accomplished by the sensory system through multisensory integration in the posterior superior temporal sulcus (pSTS). We assessed the claims regarding the involvement of the motor system in AV integration in two experiments: (i) examining the effect of articulatory suppression on the McGurk effect, and (ii) determining if motor speech regions show an AV integration profile. The hypothesis regarding experiment (i) is that if the motor system plays a role in McGurk fusion, distracting the motor system through articulatory suppression should result in a reduction of McGurk fusion. The results of experiment (i) showed that articulatory suppression results in no such reduction, suggesting that the motor system is not responsible for the McGurk effect. The hypothesis of experiment (ii) was that if the brain activation to AV speech in motor regions (such as Broca's area) reflects AV integration, the profile of activity should reflect AV integration: AV > AO (auditory-only) and AV > VO (visual-only). The results of experiment (ii) demonstrate that motor speech regions do not show this integration profile, while the pSTS does. Instead, activity in motor regions is task-dependent. The combined results suggest that AV speech integration does not rely on the motor system.

INTRODUCTION

Visible mouth movements provide information regarding the phonemic identity of auditory speech sounds and have been demonstrated to improve the perception of heard speech (Sumby & Pollack, 1954). That auditory and visual speech signals interact is further exemplified by McGurk fusion, whereby for example, auditory /pa/ overlaid onto visual /ka/ produces the percept of /ta/ (McGurk & MacDonald, 1976).

Given that the two modalities interact, researchers have attempted to determine the neural correlates of this multisensory processing. Some researchers have highlighted the importance of the posterior superior temporal sulcus (pSTS) in audiovisual (AV) speech integration in humans (Beauchamp et al., 2004a; Callan et al., 2004; Campbell et al., 2001).

The pSTS is a good candidate for multisensory integration, given its position between auditory and visual association cortex (Beauchamp et al., 2004b), and anatomical studies of the macaque brain that have shown strong anatomical connectivity of this area with different sensory cortices (Jones and Powell, 1970; Mesulam & Mufson, 1982; Yeterian & Pandya, 1985; Seltzer & Pandya, 1978; 1980). Functional magnetic resonance imaging (fMRI) studies have shown that the blood oxygenation level dependent (BOLD) response of this region is consistent with AV integration. Calvert et al. (2000), for example, found that the pSTS exhibited an increased response to a multisensory stimulus than to stimuli from the individual modalities: AV > AO (auditory-only) and AV > VO (visual-only), and Beauchamp et al. (2004a) demonstrated a patchy organization in the pSTS, with some voxels favoring unisensory stimuli and some voxels maximally sensitive to audiovisual stimuli. Additional studies have suggested a causal role for the pSTS in audiovisual integration. Beauchamp et al. (2010) localized the pSTS multisensory area in individual subjects with the conjunction of AO and VO in fMRI, then presented subjects with McGurk stimuli while simultaneously pulsing the pSTS with transcranial magnetic stimulation (TMS). Compared to baseline, subjects reported significantly fewer fusion responses, implicating this region in successful McGurk fusion. Nath and Beauchamp (2012) exploited individual differences in susceptibility to McGurk fusion; susceptibility to the illusion was positively correlated with the BOLD response in the functionally localized pSTS region (AO conjunction VO). These studies converge on the pSTS as an important region in integrating a multisensory stimulus into a unified percept.

However, neuroimaging data also show activation of motor speech regions to lipreading and AV speech (Callan et al., 2003; Calvert and Campbell, 2003; Campbell et al., 2001; MacSweeney et al., 2000; Miller and D'Esposito, 2005; Ojanen et al., 2005; Paulesu et al., 2003; Sekiyama et al., 2003; Skipper et al., 2005; 2007). Ojanen et al. (2005) presented AV vowels that were either congruent or incongruent in an fMRI study and found activation in both the STS and Broca's area (BA44 and BA45), with only Broca's area showing increased activation for conflicting stimuli. Inspired by the motor theory of speech perception (Liberman and Mattingly, 1985), in which the perceptual units of speech sounds are motor commands, the authors suggested that Broca's area performs AV integration through overlapping activation of articulatory representations by the two modalities, resulting in increased activation for incongruent stimuli as a wider range of articulatory representations activate. Skipper et al. (2007) found activation in a frontal-motor network including posterior Broca's area (BA44), dorsal premotor cortex (right), ventral premotor cortex (left), and primary motor cortex (left) using both congruent and McGurk audiovisual stimuli. BOLD timecourses to McGurk /ta/ most correlated with those of congruent /ta/ in motor regions, while timecourses in auditory and visual cortex at first correlated with the congruent stimulus for each respective modality (/pa/ for auditory, /ka/ for visual) while later correlating with congruent /ta/. The authors proposed a network for audiovisual speech in an analysis-by-synthesis framework reliant upon frontal-motor brain structures (primarily the pars opercularis of Broca's area and dorsal premotor cortex) to activate articulatory representations that constrain perception through feedback to sensory regions (Stevens and Halle, 1967).

The hypothesis that the motor system contributes to AV speech integration is further supported by TMS studies. Watkins et al. (2003) showed that motor-evoked potentials (MEPs) recorded from the lips during stimulation of face motor cortex were significantly enhanced when subjects viewed speech lip movements compared to nonspeech facial movements, while Sato et al. (2009) found that MEPs recorded from the tongue during stimulation of tongue motor cortex were significantly enhanced when perceiving tongue-related audiovisual syllables compared to lip-related syllables. These results support a somatotopic response of the motor system during the processing of audiovisual speech.

In addition to the neuroimaging data, there is behavioral evidence that supports the notion that the motor system contributes to AV speech integration. Sams et al. (2005) presented subjects with a standard McGurk paradigm, but included a condition in which subjects did not view a visual speech stimulus, but instead silently mouthed congruent or incongruent syllables along with auditory speech. They found that incongruent self-articulation (i.e., audio /pa/, articulate /ka/) produced an interference effect, with the proportion of correctly identified auditory /pa/ reduced from 68% to 33%. The authors posited that, given this effect, AV integration is driven through the activation of articulatory representations.

Given these two broad sources of evidence, Okada & Hickok (2009) hypothesized that both the pSTS and the motor system contribute to the processing of visual speech. However, while both systems may contribute to AV integration, a closer look suggests rather different roles. Ojanen et al. (2005) found that Broca's area generates more activity for incongruent than congruent AV stimuli; the same contrast revealed no activity in the pSTS. Miller and D'Esposito (2005) found more activity for AV stimuli that are perceptually unfused than for fused stimuli in the IFG, and the reverse pattern in the pSTS. Fridriksson et al. (2008) found more activity for speech videos with a reduced compared to a smooth frame rate in the motor system, including Broca's area, without seeing these effects in the pSTS. While the activity of the pSTS does show effects of perceptual fusion and stimulus synchrony (Stevenson et al., 2011), it is important to note that activations to AV speech in the motor system and the pSTS tend to dissociate such that Broca's area is more active when AV integration fails or results in conflicting cues and pSTS is more active when AV integration succeeds. This argues strongly for different roles of the two regions and hints that Broca's area may be more involved in conflict resolution as suggested in other linguistic domains, whereas the pSTS may be more involved in cross-sensory integration per se. The observation that prelinguistic infants show the McGurk effect (Rosenblum et al., 1997) is broadly consistent with this view in that it demonstrates that the ability to articulate speech is not necessary for AV integration.

These results suggest that activations in the motor system during experiments may not reflect AV integration *per se*, but something else. One alternative explanation for these activations is that the motor system responds due to demands on response selection, contingent upon the particular task in the experiment. A recent study by Venezia et al. (2012) found that during an auditory syllable discrimination task, motor speech regions showed a negative correlation with response bias, while no regions in the temporal lobe showed such a correlation. Response bias is the threshold at which subjects select one response over another, independent from perceptual analysis, suggesting that activations in

the motor system during auditory speech perception may reflect response selection rather than perceptual analysis. This finding from auditory speech perception may account for motor activations to AV speech as well.

The goal of the present study was to assess the claim that the motor system plays a primary and necessary role in AV integration, but it is also set up to assess a weaker claim that the motor system plays a secondary, modulatory role in AV integration. We refer to these hypotheses generally as “the motor hypothesis” and distinguish variants in the strength of motor involvement as needed. The motor hypothesis generates a number of predictions, including the following: (i) engaging the motor speech system with a secondary task should modulate the perception of audiovisual speech, strongly if the motor system plays a primary role and more weakly if it plays a secondary role, and (ii) motor speech regions should exhibit a physiological response that is characteristic of cross-model integration. Previous literature (Calvert et al., 2001) had emphasized the importance of supra-additivity as a defining feature of multisensory integration, requiring that the multisensory response be larger than the sum of the individual unisensory responses. However, Beauchamp (2005) suggested more relaxed criteria for identifying multisensory areas, such as requiring that the multisensory response be greater than the larger of the unisensory responses; that is, in effect, greater response to each of the individual modalities in isolation rather than summed (AV > AO and AV > VO; rather than AV > AO+VO). We used this relaxed criterion to assess whether activity in the motor system reflects AV integration in speech.

Experiment 1 was designed to assess the first prediction, that modulating articulation should modulate perception of audiovisual speech. We build on Sams et al. (2005), who found evidence in support of this prediction by demonstrating that a subject’s own articulation (/ka/) during the presentation of an auditory stimulus (/pa/) can produce McGurk-like interference. In our study, we presented subjects with a McGurk mismatch stimulus (auditory /pa/, visual /ka/) while modulating the motor system by having subjects articulate subvocally throughout stimulus presentation in a manner that should interfere with and therefore reduce the frequency of McGurk fusion, if the motor system were a critical component of the effect. To do this we chose a syllable sequence for subjects to articulate that was congruent with the auditory stimulus and incongruent with the visual stimulus in terms of place-of-articulation. Put differently, the visual signal in an AV mismatch stimulus tends to pull the percept away from the auditory signal. If this pull is mediated by the motor system, then aligning the listener’s motor articulation with the auditory signal should minimize the pull. If the pull away from the auditory signal is mediated by cross-sensory interaction between auditory and visual signals (rather than sensorimotor interaction), then motor modulation should have no effect.

Experiment 2 was designed to test the second prediction of the motor hypothesis: that motor speech regions will show an AV integration activation profile. If motor speech regions are involved in AV integration, then they should show (i) a response to both auditory and visual speech, and (ii) a larger response to multisensory speech than to auditory and visual speech in isolation (AV > A and AV > V). In an fMRI study utilizing a block design, we presented subjects with auditory-only (AO), visual-only (VO), audiovisual (AV), and McGurk mismatch speech, as well as an articulatory rehearsal condition to identify areas involved in

speech production. If the motor system contributes to AV speech integration, then motor speech areas, particularly the pars opercularis of Broca's area and premotor cortex (implicated by previous research), should show this AV integration activation profile. If the motor system does not contribute to AV speech integration, than these areas would show a profile *inconsistent* with AV integration.

EXPERIMENT 1

The objective of experiment 1 was to assess whether direct modulation of the listener's motor system via concurrent speech articulation would modulate the strength of the McGurk effect. Two versions of Experiment 1 are reported.

METHODS – Experiment 1a

Subjects—Thirteen right-handed, native speakers of English (aged 18–30 years, 11 females) volunteered for participation. Subjects had normal or corrected to normal vision, and no hearing impairment. Subjects were given course credit for their participation. Consent was acquired from each subject before participation in the study, and all procedures were approved by the Institutional Review Board of UC Irvine.

Stimuli—Auditory stimuli (AO) consisted of recordings of a native speaker of English producing the speech sounds /pa/, ta/ and /ka/. The duration of each recording was 1000ms, and the duration of the auditory speech was ~300ms for each syllable, digitized at 44,100 Hz. Each stimulus consisted of 4 repetitions of the same syllable. We presented subjects with 4 repetitions because we wanted to ensure that articulation had the maximal opportunity to impact perception. Low-amplitude continuous white noise (level set to 10% RMS of speech) was added to each stimulus to ensure McGurk fusion as well as mask any sounds inadvertently produced during suppression. We created video recordings of the same speaker articulating the same speech sounds at a frame rate of 30 fps. Congruent audiovisual stimuli (AV) were generated by overlaying the auditory stimuli onto the corresponding visual stimuli and aligning the onset of the consonant burst with the audio captured in the video recordings. In addition, one mismatch video was generated by overlaying auditory /pa/ onto visual /ka/ (McGurk-inducing). During stimulus presentation, stimulus loudness was set at a level that was clearly audible and comfortable for each subject.

Procedure—Subjects were informed that they would be viewing videos of a speaker articulating syllables, and were asked to make decisions regarding the identity of the acoustic stimuli in a 3AFC design among /pa/, /ta/, and /ka/. Specifically, they were instructed to "report the sound that they heard". Subjects made these judgments while simultaneously performing a secondary task, adapted from Baddeley (1981), which consisted either of continuously articulating the sequence "/pa/…/ba/" without producing sound or continuously performing a finger-tapping sequence, 1-2-3-4-5-5-4-3-2-1 (1 = thumb, 5 = pinky). For the suppression task, we chose the sounds /pa/ and /ba/ because /pa/ is identical to the auditory portion of our mismatch stimulus, and /ba/ differs only on the feature of voicing, or the onset time of the vibration of the vocal folds. Otherwise, the vocal tract configuration during the articulation of these two consonants above the larynx is

identical. If visual speech influences heard speech via activation of motor representations, then saturating the motor system with auditory-congruent representations should strengthen activation in favor of the auditory stimulus, lessening the effect of the incongruent visual stimulus. Subjects were instructed to perform both tasks at 2Hz, and were cued at that rate by an onscreen flickering fixation point that disappeared during stimulus presentation. Subjects were instructed to continuously perform the task throughout stimulus presentation. Subjects performed the same task (articulation or finger tapping) throughout a given experimental run.

Stimuli were blocked by modality (AO, AV) and task (articulatory suppression, finger-tapping) in 4 experimental runs. Mismatch stimuli were presented during AV runs. AV runs were presented first to prevent subjects from guessing the incongruent nature of the mismatch stimuli. Order of task was counterbalanced across subjects. Ten trials of each stimulus were presented in random order in each run. Subjects made their responses by indicating their decision on answer sheets provided to them. Once the subject completed a trial, she or he cued the onset of the next trial in a self-paced fashion. Each trial began by cueing the subject to get ready, and then began with a button press when the subject was ready to begin the secondary task. The task cue flickered for 4 seconds, at which point the stimulus was presented followed by a prompt to respond. Stimuli were delivered through a laptop computer with Matlab software (Mathworks, Inc, USA) utilizing Psychtoolbox (Brainard, 1997; Pelli, 1997) and headphones (Sennheiser HD280).

To determine whether there was a McGurk effect, we compared performance on auditory identification of the AO /pa/ stimulus to the mismatch stimulus, with the expectation that successful interference results in reduced auditory identification performance in the mismatch condition. Therefore, we analyzed the data in a 2x2 design, crossing stimulus (AO /pa/, mismatch) x task (finger-tapping, articulatory suppression) in order to determine if task had an effect on the strength of the McGurk effect. Even though only AO /pa/ and mismatch trials were included in the analysis, we included the other stimuli in the experiment so that subjects gave a range of responses throughout the experiment in order to prevent them from guessing the nature of the mismatch stimulus.

RESULTS - Experiment 1a

The average correct identification of the auditory-only and *congruent* AV stimuli was at or near ceiling across /pa/, /ta/, and /ka/ for both secondary tasks. Fig. 1 illustrates that the McGurk effect was equally robust during both secondary tasks. Given the presence of ceiling and floor effects, we performed non-parametric statistical examinations of the data (Kruskal-Wallis) that are less sensitive to outliers than parametric tests. Subjects reported significantly more /pa/ responses during the auditory-only /pa/ condition than the Mismatch condition, an effect of condition, $\chi^2(1, N=52) = 40.94$, $p < 0.001$, indicating a successful McGurk effect. There was no effect of task on /pa/ responses in the Mismatch condition, $\chi^2(1, N=26) = 0.131$, $p = 0.718$, nor was there an effect of task on /ta/ (fusion) responses in the Mismatch condition, $\chi^2(1, N=26) = 1.783$, $p = 0.182$, indicating no effect of articulatory suppression on the McGurk effect. All reported Kruskal-Wallis tests are Bonferroni corrected for multiple comparisons with a familywise error rate of $p < 0.05$ (per-comparison

error rates of $p < 0.0167$). The majority of responses in the mismatch condition during articulatory suppression were fusion responses (/ta/, 95%), rather than the visual capture response (/ka/, 2%). Consistent with a cross-sensory model of the source of AV integration, and against the predictions of the motor hypothesis, these results strongly suggest that articulatory suppression does not affect McGurk fusion.

DISCUSSION - Experiment 1a

Subjects correctly identified auditory-only and congruent audiovisual syllables, but performance changed dramatically during perception of the incongruent stimuli. This is a classic McGurk effect (McGurk & MacDonald, 1976).

Against the predictions of the motor hypothesis, we did not see any difference between subjects' responses during the articulatory suppression task and the finger-tapping task. In a framework that highlights the importance of articulatory representations in integrating AV speech, one would expect any distracting articulation to reduce McGurk fusion. In our experiment, subjects' own articulations were *congruent* with the auditory stimulus, which should have the strongest possible effect. Instead, the articulatory suppression task showed no effect. This suggests that the McGurk effect is not mediated or even modulated by the motor system.

One possible issue with our results is that subjects may have failed to articulate simultaneously with the auditory stimulus. This is unlikely given that subjects were cued to begin articulation before the onset of the stimulus and continue throughout its duration at a fairly rapid rate (2 Hz), and because subjects fused at nearly 100% during the articulation task, implying that this would have had to happen on nearly every trial. In a previous version of the experiment that was run as a pilot study, we used a single stimulus presentation with a single simultaneous articulation, in accordance with Sams et al., 2005, and did not observe the reported interference effect. This led us to adopt the current design with 4 stimulus repetitions and rapid articulatory suppression in an attempt to afford the motor system the most chance to influence perception. However, one might still argue that the motor system was not sufficiently driven by this task. A second issue concerns our usage of /pa/ and /ba/ during the articulatory suppression task. It is possible that the use of more than one syllable caused some form of confusion and led subjects to rely more on the visual stimulus, contaminating the results. A third potential issue with our design is the presence of 4 stimulus repetitions, which may have somehow altered the results due to subjects making a collective judgment on multiple stimuli rather than a single stimulus. To address these concerns, we ran a second experiment in which we employed a rapid articulatory suppression of /pa/ alone without cueing (i.e., as fast as possible), and trials that consisted of only a single stimulus presentation.

METHODS – Experiment 1b

Subjects—Seventeen right-handed, native speakers of English (aged 18–39 years, mean 21 years, 10 females) volunteered for participation. Subjects had normal or corrected to normal vision, and no hearing impairment. Subjects were given course credit for their participation.

Consent was acquired from each subject before participation in the study, and all procedures were approved by the Institutional Review Board of UC Irvine.

Stimuli—Stimuli were identical to experiment 1a, with the following modifications: the duration of each stimulus was lengthened to 2000ms (syllable duration the same), white noise level was increased to 20% RMS of speech, and each stimulus consisted of only a single presentation.

Procedure—The experimental procedure was identical to experiment 1b, with the following modifications. We altered the articulatory suppression task such that subjects were instructed to articulate /pa/ silently and as rapidly as possible from when the trial began and throughout stimulus presentation, instead of cued to articulate /pa/…/ba/ at 2Hz. We replaced the finger-tapping task with a baseline condition with no secondary task. Stimuli were blocked by modality (AO, AV) and condition (baseline, articulatory suppression) in 4 experimental runs. Order of condition and modality was different (partially counterbalanced) for each subject. Ten trials of each stimulus were presented in random order in each run. Subjects made their responses by pressing the appropriate key on the keyboard. Once the subject completed a trial, she or he cued the onset of the next trial in a self-paced fashion. Each trial began by cueing the subject to get ready, and then began with a button press when the subject was ready to begin the trial. A fixation 'x' appeared for 1.5s, followed by the stimulus.

The data were analyzed in the same manner as experiment 1a, replacing the finger-tapping condition with the baseline condition.

RESULTS - Experiment 1b

The results are consistent with experiment 1a. Average correct identification of the auditory-only and *congruent* AV stimuli was at ceiling across /pa/, /ta/, and /ka/ for both secondary tasks. Fig. 2 illustrates that the McGurk effect was equally robust during baseline and articulatory suppression. As in experiment 1a, we performed non-parametric statistical examinations of the data (Kruskal-Wallis). Subjects reported significantly more /pa/ responses during the auditory-only /pa/ condition than the Mismatch condition, an effect of condition, $\chi^2(1, N=68) = 45.80$, $p < 0.001$, indicating a successful McGurk effect. There was no effect of task on /pa/ responses in the Mismatch condition, $\chi^2(1, N=34) = 0.30$, $p = 0.584$, nor was there an effect of task on /ta/ (fusion) responses in the Mismatch condition, $\chi^2(1, N=34) = 0.09$, $p = 0.762$, indicating no effect of articulatory suppression on the McGurk effect. All reported Kruskal-Wallis tests are Bonferroni corrected for multiple comparisons with a familywise error rate of $p < 0.05$ (per-comparison error rates of $p < 0.0167$). There were some differences from experiment 1a in the response rate for each alternative, but they were qualitatively similar, with a majority fusion responses (/ta/, 68%).

DISCUSSION - Experiment 1b

As in experiment 1a, subjects correctly identified auditory-only and congruent audiovisual syllables, but performance changed dramatically during perception of the incongruent stimuli, confirming the presence of a McGurk effect (McGurk & MacDonald, 1976) under

conditions of articulatory suppression. There were some differences in the overall fusion rate between the two experiments (~90% in experiment 1a and ~65% in 1b), and concomitant differences in visual capture and auditory perceptions. The difference in fusion rates may be largely explained by the difference in presentation: 4 repetitions of the same stimulus were used in experiment 1a, while only a single presentation used in 1b. In addition, the noise level increase may have affected some subjects' judgments. However, the alterations in the experimental design did not qualitatively change the results: McGurk fusion rate does not change from baseline during articulatory suppression. This allays the noted concerns from experiment 1a.

The result that the McGurk effect is not weakened under articulatory suppression conflicts with the results of Sams et al. (2005). However, the discrepancy can be explained by closely examining their results. Considering the type of response in their study (fusion /ta/ or visual/articulatory capture /ka/), the proportion of fusion responses was the same in the baseline condition as during their articulation condition (23%), with only "capture responses" (percept is congruent with the visual stimulus) increasing with articulation (46% vs. 9%). This trend held for all of their experimental conditions, including for a written /ka/ (26% /ka/). This is different from most McGurk paradigms, in which the bulk of the interference effect derives from fusion rather than visual capture. The interference effect of the written stimulus, along with their effects being driven by capture rather than fusion, may be partially explained by the high-amplitude noise added to the auditory stimulus in order to drive baseline /pa/ identification down to ~68%. In this light, the interference effect that obtained from this study is relatively weak, and may have resulted from response bias induced by the noisy auditory stimulus.

In summary, we found no evidence that behavioral interference involving the motor speech system modulates the McGurk effect, casting doubt on both strong and weak versions of the motor hypothesis of AV integration.

EXPERIMENT 2

The goal of experiment 2 was to use fMRI to examine the profile of activation in motor speech regions in response to auditory, visual, and audiovisual speech and to compare this profile with the one observed in the superior temporal sulcus. We were particularly interested in determining whether speech motor areas (specifically, the pars opercularis of Broca's area and premotor cortex) exhibit an AV integration profile (AV > AO and AV > VO).

METHODS – Experiment 2

Subjects—Twenty right-handed, native speakers of English (aged 20–30 years, 8 males) volunteered for participation. Subjects had normal or corrected-to-normal vision and no hearing impairment, and reported no history of neurological disorder. Subjects were paid $30 an hour for their participation. Consent was acquired from each subject before participation in the study, and all procedures were approved by the Institutional Review Board of UC Irvine.

Stimuli and Design—The stimuli from experiment 2 were identical to experiment 1, except for the following: all stimuli had duration 1000ms and the noise level was set to 25% RMS of speech. Visual-only (VO) stimuli were added, consisting of the same videos as the congruent AV stimuli with no sound. In addition, an articulatory rehearsal condition (ART) was added, cued by a flickering fixation cross. In sum, the experiment consisted of a 3x3 design, condition (AO, VO, AV) x stimulus (/pa/, /ta/, /ka/), plus two additional conditions, mismatch (MM) and ART.

Procedure—Subjects were informed that they would view videos of a talker articulating the speech sounds /pa/, /ta/, and /ka/ and instructed to make decisions regarding the identity of the stimuli. Trials consisted of a block of 10 sequential identical speech sounds followed by 2.5s of fixation. Subjects were instructed to pay attention throughout the duration of the trial, and at the end of the block to identify the speech sound in audio and audiovisual trials and the intended speech sound in visual trials. As in the behavioral experiment, subjects were not informed of the incongruent nature of the mismatch stimulus, although two subjects were aware of the presence of a mismatch stimulus. Responses were made with a response box using the left hand. Subjects assigned a distinct button for each possibility in a 3AFC design among /pa/, /ta/, and /ka/, with three fingers assigned to the respective buttons. Subjects were instructed to make their response within 2s of stimulus offset. AO trials were presented alongside a still image of the speaker's face, while VO trials were presented in silence (aside from the background scanner noise). During ART trials, the cue to articulate was a fixation cross that flickered at 2 Hz, and subjects were instructed to produce the sequence /pa/…/ta/…/ka/ repeatedly throughout the duration of flickering (10s) without producing sound or opening their mouth while still making movements internal to the vocal tract including tongue movements. Subjects stopped articulating when the fixation cross stopped flickering. Stimuli were delivered with Matlab software (Mathworks, Inc, USA) utilizing Cogent (http://vislab.ucl.ac.uk/cogent_2000.php) and MR compatible headphones. The experiment consisted of 9 runs – 1 practice run, 6 functional runs, 2 localizer runs - and 1 anatomical scan. The practice run was utilized to familiarize subjects with the stimuli and task, and no data were analyzed from this run. Four trials of each condition along with 4 rest trials (still image of speaker's face) were presented in random order within each functional run (24 trials total). Due to a coding error, two subjects were given slightly uneven amounts of trials from the AO and VO conditions, with one of those subjects also given slightly fewer mismatch trials. The localizer runs consisted solely of VO and rest trials, in order to obtain functionally independent ROIs for further analysis (12 VO, 6 rest per run). The stimuli and task remained the same throughout these two localizer runs. Following this, we collected a high-resolution anatomical scan. In all, the subjects were in the scanner less than an hour.

fMRI Data Collection and Preprocessing—MR images were obtained in a Philips Achieva 3T (Philips Medical Systems, Andover, MA) fitted with an eight channel RF receiver head coil at the high field scanning facility at UC Irvine. We first collected a total of 1110 T2*-weighted EPI volumes over 9 runs using Fast Echo EPI in ascending order (TR=2.5s, TE=25ms, flip angle = 90°, in-plane resolution = 1.95mm × 1.95mm, slice thickness = 3mm with 0.5mm gap). The first four volumes of each run were collected before stimulus presentation and discarded to control for saturation effects. After the functional

scans, a high-resolution T1-weighted anatomical image was acquired in the axial plane (TR=8ms, TE=3.7ms, flip angle=8°, size=1mm isotropic).

Slice-timing correction, motion correction and spatial smoothing were performed using AFNI software (http://afni.nimh.nih.gov/afni). Motion correction was achieved by using a 6-parameter rigid-body transformation, with each functional volume in a run first aligned to a single volume in that run. Functional volumes were aligned to the anatomical image, and subsequently aligned to Talairach space (Talairach and Tournoux, 1988). Functional images were resampled to 2.5mm isotropic voxels, and spatially smoothed using a Gaussian kernel of 6mm FWHM.

First-level analyses were performed on each individual subject's data using AFNI's 3dDeconvolve function. The regression analysis was performed to find parameter estimates that best explained variability in the data. Each predictor variable representing the time course of stimulus presentation was convolved with the hemodynamic response function and entered into the general linear model. The following five regressors of interest were used in the experimental analysis: auditory-only speech (AO), visual-only speech (VO), congruent audiovisual speech (AV), mismatch (MM), and articulation (ART). The six motion parameters were included as regressors of no interest. The independent localizer data were analyzed in the same fashion, with the single regressor of interest, the VO condition. A second-level analysis was then performed on the parameter estimates, using AFNI's 3dANOVA2 function. Using an FDR correction for multiple comparisons, a threshold of $q < 0.05$ was used to examine activity above baseline for each condition and for the following contrasts: [AV > AO], [AV > VO], and [MM > AV].

In order to compare the profile of activation in motor areas to those of the pSTS, we split our functional data into even and odd runs. The even runs were used to localize the pSTS multisensory region using a conjunction analysis of AO and VO (Nath & Beauchamp, 2012) with individual uncorrected $p < 0.05$ for each condition. We justify the use of this liberal threshold because a) the threshold was used only to select the ROIs, b) the conjunction analysis produces a combined statistical threshold much more stringent than the individual thresholds, and c) the power is greatly reduced because the data are split between localization and analysis. The conjunction analysis from the even runs resulted in a pSTS ROI from both hemispheres. The data from odd runs were averaged within each ROI and the means entered into *t*-tests. No motor speech areas were localized using the conjunction analysis, so we used the results of the localizer analysis (VO > rest, $q < 0.01$) to define two frontal-motor ROIs previously reported to be engaged in AV speech integration, posterior Broca's area (pars opercularis) and dorsal premotor cortex of the precentral gyrus (Ojanen et al., 2004; Skipper et al., 2007). The parameter estimates for each subject for each condition from the functional runs were averaged within each ROI and the means entered into a statistical analysis.

RESULTS - Experiment 2

Behavioral performance—Subjects accurately identified 93% of stimuli in the AO condition, 68% in the VO condition, 97% in the AV condition, and 4% in the mismatch condition (86% fusion, 8% visual capture). Analyzing behavioral performance during the

AV, AO, and VO conditions showed a significant main effect of condition, $F(2,38) = 74.607$, $p < 0.001$, a significant main effect of syllable, $F(2,38) = 24.239$, $p < 0.001$, and a significant interaction, $F(4,76) = 41.229$, $p < 0.001$. There was a significant effect of syllable in the AO condition, $F(2,38) = 6.766$, $p = 0.003$, no effect of syllable in the AV condition, $F(2,38) = 1.401$, $p = 0.259$, and a significant effect of syllable in the VO condition, $F(2,38) = 63.004$, $p < 0.001$.

Individual comparisons in the AO condition revealed that identification of AO /pa/ was lower than AO /ta/, $t(19) = 2.613$, $p = 0.017$ (two-tailed), AO /pa/ was lower than AO /ka/, $t(19) = 2.626$, $p = 0.017$ (two-tailed), with no difference between AO /ta/ and AO /ka/, $t(19) = 0.567$, $p = 0.577$ (two-tailed). Individual comparisons in the VO condition revealed that identification of VO /pa/ was greater than VO /ta/, $t(19) = 3.796$, $p = 0.001$ (two-tailed), VO /pa/ was greater than VO /ka/, $t(19) = 14.312$, $p < 0.001$ (two-tailed), and VO /ta/ was greater than VO /ka/, $t(19) = 5.998$, $p < 0.001$ (two-tailed). All reported t-tests are Bonferroni corrected for multiple comparisons with a familywise error rate of $p < 0.05$, with per-comparison error rates of $p < 0.0167$.

We suspected that the poor performance in the VO condition was due to the similarity between /ta/ and /ka/. The difference in these two stimuli resides in the place of articulation of the tongue, which is difficult to see when viewing the face. The visual similarity of these two stimuli suggests that the consonants belong to the same viseme (the visual counterpart to the phoneme; Fisher, 1968). On the contrary, bilabial /pa/ is easily discriminated from /ta/ and /ka/ due to the involvement of lip closure. However, our statistical examination of percent correct indicated that /ta/ was significantly more accurate than /ka/. This suggested that subjects, when faced with ambiguous /ta/ or /ka/, were biased to respond /ta/, resulting in higher accuracy than chance in the /ta/ condition and lower accuracy than chance in the /ka/ condition. Thus, we decided to examine these data using signal detection theory, which allowed us to account for response bias and obtain a true measure of subjects' ability to discriminate these two stimuli. We analyzed only the /ta/ and /ka/ data using a standard 2AFC calculation of d' (Macmillan and Creelman, 2005), treating /ta/ responses to /ta/ as hits and /ta/ responses to /ka/ as false alarms. We are justified in excluding /pa/ from the decision space as subjects falsely identified VO /pa/ as /ta/ or /ka/ 0% of the time, and falsely identified VO /ta/ and /ka/ as /pa/ 0% of the time, indicating that subjects never considered /pa/ as a possibility during identification of /ta/ and /ka/. Our results showed a d' of 0.11, indicating that subjects were effectively at chance discriminating VO /ta/ and /ka/, confirming our expectation.

fMRI Analyses—Activation relative to "rest" (still image of speaker's face; no auditory stimulus) for each condition is shown in fig. 3. AV generated greater activity than AO in lateral occipital lobe bilaterally, right hemisphere IFG, left premotor cortex, and right parietal lobe (fig 4; table 1). AV generated greater activity than VO in the superior temporal lobe bilaterally and throughout the default network (Buckner et al., 2008; fig. 4; table 2). MM activated the motor speech network significantly more than AV speech, including the pars opercularis of Broca's area, anterior insula and left premotor cortex (fig. 4; table 3). The VO localizer activated a similar set of brain regions as the VO condition in the experiment

(fig. 3); we selected the left and right hemisphere pars opercularis and dorsal precentral gyrus activations as our ROIs for further analysis.

Fig. 5, top, illustrates the results of the ROI analyses in these left hemisphere motor areas localized through the independent VO runs. Both regions in the left hemisphere were strongly activated by the ART condition, confirming that they were indeed motor speech areas. In the right hemisphere, only the premotor ROIs were activated by the ART condition. All ROIs were strongly activated by the VO condition. However, comparisons among the conditions in both regions revealed an activation profile inconsistent with AV integration. AO produced little activation, and VO was significantly greater than AV and the mismatch condition. By contrast, fig. 6 illustrates the results of the analyses in pSTS. Bilateral pSTS, localized through AO conjunction VO using independent runs, exhibited the expected AV integration pattern, with AV conditions producing more than either AO or VO alone. We were unable to effectively localize motor regions using this conjunction analysis.

One could argue that using a unimodal localizer (VO) to select the ROIs biased the analysis against finding an AV integration region in motor cortex. To rule out this possibility, we redefined the ROIs in the pars opercularis and premotor cortex using AV vs. rest from the even functional runs and reran the analysis on the odd functional runs. The results from this analysis are therefore biased *in favor* of AV integration. Regardless, the same response profile resulted from the analysis as indicated by fig. 5, bottom.

Given the disparity in behavioral performance for individual stimuli during the VO condition (/pa/ at ceiling, /ta/ and /ka/ at chance), we decided to perform a post-hoc analysis to determine if activation in the motor system during VO followed this pattern. We re-ran the individual subject deconvolution analyses, replacing the regressor of interest for the VO condition with 3 individual regressors for /pa/, /ta/ and /ka/. The parameter estimates for each subject for each stimulus were averaged within both frontal ROIs obtained from the VO localizer, and the means entered into *t*-tests. Consistent with our prediction, Fig. 7 shows that activity in both the pars opercularis of Broca's area and dorsal premotor cortex was higher for /ta/ and /ka/ than for /pa/, with no difference between /ta/ and /ka/, confirming that activity was lowest when subjects were at ceiling (/pa/), and highest when at chance (/ta/, /ka/).

DISCUSSION - Experiment 2

Consistent with previous research, the whole-brain contrasts AV > AO and AV > VO each resulted in activations in the vicinity of the pSTS. The ROI analysis confirmed that this region displayed the expected AV integration profile: AV > AO and AV > VO. While the whole-brain contrast AV > AO resulted in activity in the posterior left IFG (Broca's area), the contrast AV > VO did not. The ROI analysis of posterior left IFG and left dorsal premotor cortex, two motor speech regions implicated in AV integration (Skipper et al., 2007), confirmed that the motor system did not display the AV integration profile: VO speech activated these areas significantly more than AV speech (VO > AV). Post-hoc analysis of the individual stimuli from the VO condition revealed that the ambiguous stimuli /ta/ and /ka/ drove much of the activation in these areas. These results suggest that AV integration for speech involves the pSTS but not the speech motor system.

GENERAL DISCUSSION

Neither experiment found evidence that the motor speech system is involved in AV integration, even in a weak modulatory capacity. Using behavioral measures, Experiment 1 found that strongly modulating the activity of the motor speech system via articulatory suppression did not correspondingly modulate the strength of the McGurk effect; in fact it had no effect. If the motor speech system mediates the AV integration processes that underlie the McGurk effect then we should have seen a significant of motor modulation on McGurk fusion, yet we did not. Using fMRI, Experiment 2 found that the response of motor speech areas did not show the characteristic signature of AV integration (AV > AO and AV > VO). Instead, AV stimuli activated motor speech areas significantly *less* than visual speech stimuli alone. Consistent with previous reports, response properties of the superior temporal sulcus were more in line with a region critically involved in AV integration (AV > AO and AV > VO). Taken together, these studies substantially weaken the position that motor speech areas play a significant role in audiovisual speech integration and strengthen the view that the STS is the critical site.

If motor speech areas are not involved in AV integration, why do these regions activate under some speech-related conditions, such as visual-only speech? One view is that motor speech circuits are needed for perception and therefore recruited for perceptual purposes under noisy or ambiguous conditions. There is no doubt that motor regions are indeed more active when the perceptual signal is degraded, as shown by previous studies (Miller & D'Esposito, 2005; Fridrikkson, 2009). This was evident in the present study with partially ambiguous visual-only speech generating more motor-related activity than relatively perceptible audiovisual speech. But to say that the motor system is recruited under demanding perceptual conditions only restates the facts. The critical question is: does the motor system actually aid or improve perception in any way?

A recent fMRI study suggests that this is not the case. Venezia et al. held perceptibility (d′) constant in an auditory CV syllable discrimination task and varied performance by manipulating response bias using different ratios of same to different trials (Venezia et al., 2012). Neural activity in motor speech areas was significantly negatively correlated with behaviorally measured response bias, even though perceptual discriminability was held constant. This suggests that while motor regions are recruited under some task conditions, their involvement does not necessarily result in better perceptual performance. Similar results were obtained in a purely behavioral experiment in which use-induced motor plasticity of speech articulators modulated bias but not discriminability of auditory syllables (Sato et al., 2011). The results are consistent with the motor system interacting with subject responses, but not aiding in perception, as d' would be expected to vary if this were the case.

One alternative interpretation of motor activity during speech tasks is that it is purely epiphenomenal, deriving from associations between auditory and motor speech systems that are critical for speech production, but not for speech perception. Existing models of speech motor control provide a mechanism for such an association in the form of feedback control architectures (Hickok, et al. 2011; Hickok 2012; Tourville & Guenther, 2011; Houde et al.,

2002). However, this view fails to explain why the motor system is more active in some conditions than others. If pure association were the only mechanism driving the motor activations, one would expect equal activation under all conditions; clearly, the activity differs among modalities (AO, VO, AV), stimulus quality, and task. In addition, the findings of Venezia et al. (2012) and Sato et al. (2011) point toward some role of the motor system in generating subject's responses, as response bias correlates with activity in the motor system. Since epiphenomenal, that is pure association-related activation alone cannot account for these effects, another mechanism must be driving the activations.

A third possibility is that the motor system somehow participates in response selection. As the response selection demands increase, so does activity in motor speech systems. This accounts for the correlation between response bias and motor speech region activity as well as the tendency for these regions to be more active when the perceptual stimuli are degraded and ambiguous (thus increasing the load on response selection). Previous work has implicated motor-related regions in the inferior frontal gyrus in response selection (Novick et al., 2005; Snyder et al., 2007; Thompson-Schill et al., 1997), which is broadly consistent with this view. What remains unclear is the role that lower-level motor speech areas (including the dorsal premotor cortex) play in response selection. One possibility is that they contribute via their involvement in phonological working memory, which clearly has an articulatory component (Buchsbaum 2005; Hickok et al., 2003). During syllable identification or discrimination tasks, subjects may utilize verbal working memory resources in difficult processing environments, resulting in activation in the motor system. What is clear from the evidence is that the activation of these regions does not track with speech perception or audiovisual integration. More work is needed to determine precisely the role that these motor regions play in response selection during speech comprehension.

CONCLUSIONS

The results of our experiments suggest that the motor system does not play a role in audiovisual speech integration. First, articulatory rehearsal had no effect on the McGurk effect, showing that activation of articulatory representations does not inhibit McGurk fusion, suggesting that the motor speech network and the AV integration network do not interact during McGurk fusion. Second, motor speech regions (including the pars opercularis of Broca's area and dorsal premotor cortex) exhibited an activation profile inconsistent with AV integration. Demands on response selection likely account for much of the activity in these regions during speech perception, unisensory or multisensory. Alternatively, the pSTS does exhibit such an integration pattern, consistent with previous accounts of its role in audiovisual integration.

Acknowledgments

The authors would like to thank J. Venezia for useful comments and suggestions throughout this investigation. This investigation was supported by a grant (DC009659) from the US National Institutes of Health.

References

Arnal LH, Morillon B, Kell CA, Giraud AL. Dual Neural Routing of Visual Facilitation in Speech Processing. Journal of Neuroscience. 2009; 29(43):13445–13453. [PubMed: 19864557]

Baddeley A, Eldridge M, Lewis V. The Role of Subvocalization in Reading. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology. 1981; 33(v):439–454.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nature Neuroscience. 2004; 7(11): 1190–1192. [PubMed: 15475952]

Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 2004; 41(5):809–823. [PubMed: 15003179]

Beauchamp MS, Nath AR, Pasalar S. fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. Journal of Neuroscience. 2010; 30(7):2414–2417. [PubMed: 20164324]

Beauchamp MS. Statistical criteria in fMRI studies of multisensory integration. Neuroinformatics. 2005; 3(2):93–113. [PubMed: 15988040]

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Buchsbaum BR, Olsen RK, Koch P, Berman KF. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron. 2005; 48(4):687–697. [PubMed: 16301183]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network. Annals of the New York Academy of Sciences. 2008; 1124(1):1–38. [PubMed: 18400922]

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport. 2003; 14(17):2213–2218. [PubMed: 14625450]

Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. Journal of Cognitive Neuroscience. 2004; 16(5):805–816. [PubMed: 15200708]

Calvert G, Brammer M, Campbell R. Cortical subtrates of seeing speech: still and moving faces. Neuroimage. 2001; 13(6):S513–S513.

Calvert GA, Campbell R. Reading speech from still and moving faces: The neural substrates of visible speech. Journal of Cognitive Neuroscience. 2003; 15(1):57–70. [PubMed: 12590843]

Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, … David AS. 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). Cognitive Brain Research. 2001; 12(2):233–243. [PubMed: 11587893]

Fridriksson J, Moss J, Davis B, Baylis GC, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. Neuroimage. 2008; 41(2):605–613. [PubMed: 18396063]

Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron. 2011; 69(3):407–422. [PubMed: 21315253]

Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci. 2012; 13(2):135–145. [PubMed: 22218206]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15(5): 673–682. [PubMed: 12965041]

Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. 2011; 5:82. [PubMed: 22046152]

Jones EG, Powell TPS. An anatomical study of converging sensory pathways in the cerebral cortex of the monkey. Brain Behavior and Evolution. 1970; 93(4):793–820.

Kleiner, M.; Brainard, D.; Pelli, D. Perception 36 ECVP Abstract Supplement. 2007. What's new in Psychtoolbox-3?.

Macmillan, NA.; Creelman, CD. Detection Theory: A User's Guide. Mahwah, NJ: Erlbaum; 2005.

MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, … Brammer MJ. Silent speechreading in the absence of scanner noise: an event-related fMRI study. Neuroreport. 2000; 11(8):1729–1733. [PubMed: 10852233]

McGurk H, Macdonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. [PubMed: 1012311]

Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. Journal of Neuroscience. 2005; 25(25):5884–5893. [PubMed: 15976077]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. 2012; 59(1):781–787. [PubMed: 21787869]

Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. Cognitive Affective & Behavioral Neuroscience. 2005; 5(3):263–281.

Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M. Processing of audiovisual speech in Broca's area. Neuroimage. 2005; 25(2):333–338. [PubMed: 15784412]

Okada K, Hickok G. Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. Neuroscience Letters. 2009; 452(3):219–223. [PubMed: 19348727]

Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, … Fazio F. A functional-anatomical model for lipreading. Journal of Neurophysiology. 2003; 90(3):2005–2013. [PubMed: 12750414]

Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442. [PubMed: 9176953]

Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception & Psychophysics. 1997; 59(3):347–357. [PubMed: 9136265]

Sams M, Mottonen R, Sihvonen T. Seeing and hearing others and oneself talk. Cognitive Brain Research. 2005; 23(2–3):429–435. [PubMed: 15820649]

Sato M, Buccino G, Gentilucci M, Cattaneo L. On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. Speech Communication. 2010; 52(6):533–541.

Sato M, Grabski K, Glenberg AM, Brisebois A, Basirat A, Menard L, Cattaneo L. Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. Cortex. 2011; 47(8):1001–1003. [PubMed: 21501836]

Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neuroscience Research. 2003; 47(3):277–287. [PubMed: 14568109]

Seltzer B, Pandya DN. Afferent cortical connections and architectonics of superior temporal sulcus and surrounding cortex in rhesus-monkey. Brain Research. 1978; 149(1):1–24. [PubMed: 418850]

Seltzer B, Pandya DN. Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus-monkey. Brain Research. 1980; 192(2):339–351. [PubMed: 6769545]

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage. 2005; 25(1):76–89. [PubMed: 15734345]

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex. 2007; 17(10):2387–2399. [PubMed: 17218482]

Snyder HR, Feigenson K, Thompson-Schill SL. Prefrontal cortical response to conflict during semantic and phonological tasks. Journal of Cognitive Neuroscience. 2007; 19(5):761–775. [PubMed: 17488203]

Stevens, KN.; Halle, M. Remarks on analysis by synthesis and distinctive features. In: Wathen-Dunn, W., editor. Models for the perception of speech and visual form. Caimbridge, MA: MIT Press; 1967. p. 88-102.

Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. Neuroimage. 2011; 55(3):1339–1345. [PubMed: 21195198]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America. 1954; 26(2):212–215.

Talairach, J.; Tournoux, P. Co-planar Stereotaxic Altas of the Human Brain. New York: Thieme Medical Publishers; 1988. p. 122

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. Proceedings of the National Academy of Sciences USA. 1997; 94(26):14792–14797.

Tourville JA, Guenther FH. The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes. 2011; 26(7):952–981. [PubMed: 23667281]

Venezia JH, Saberi K, Chubb C, Hickok G. Response Bias Modulates the Speech Motor System during Syllable Discrimination. Frontiers in Psychology. 2012; 3:157. [PubMed: 22723787]

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003; 41(8):989–994. [PubMed: 12667534]

Yeterian EH, Pandya DN. Corticothalamic connections of the posterior parietal cortex in the rhesus-monkey. Journal of Comparative Neurology. 1985; 237(3):408–426. [PubMed: 4044894]

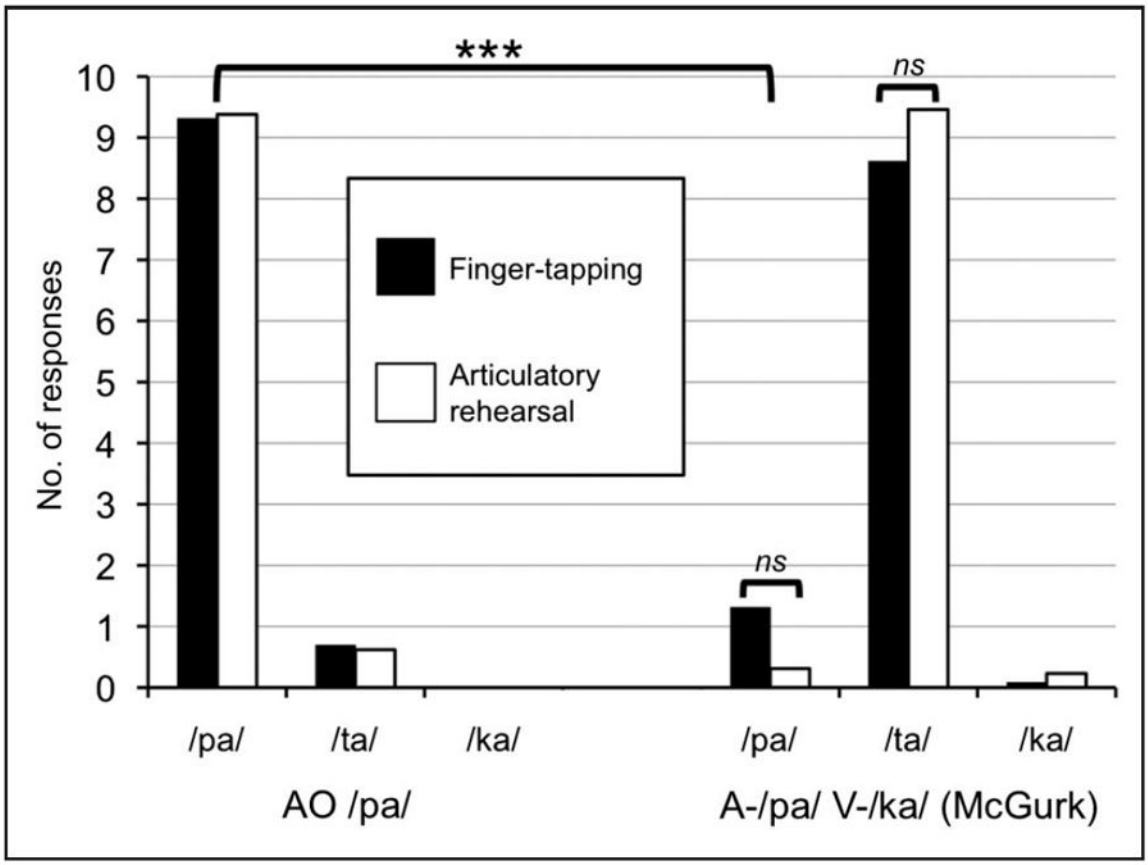

Figure 1.

Average number of responses for each alternative during Experiment la for the AO /pa/ and MM (A-/pa/, V-/ka/; McGurk) conditions during the finger-tapping task (black bars) and the articulatory suppression task (white bars). *ns* = not significant.

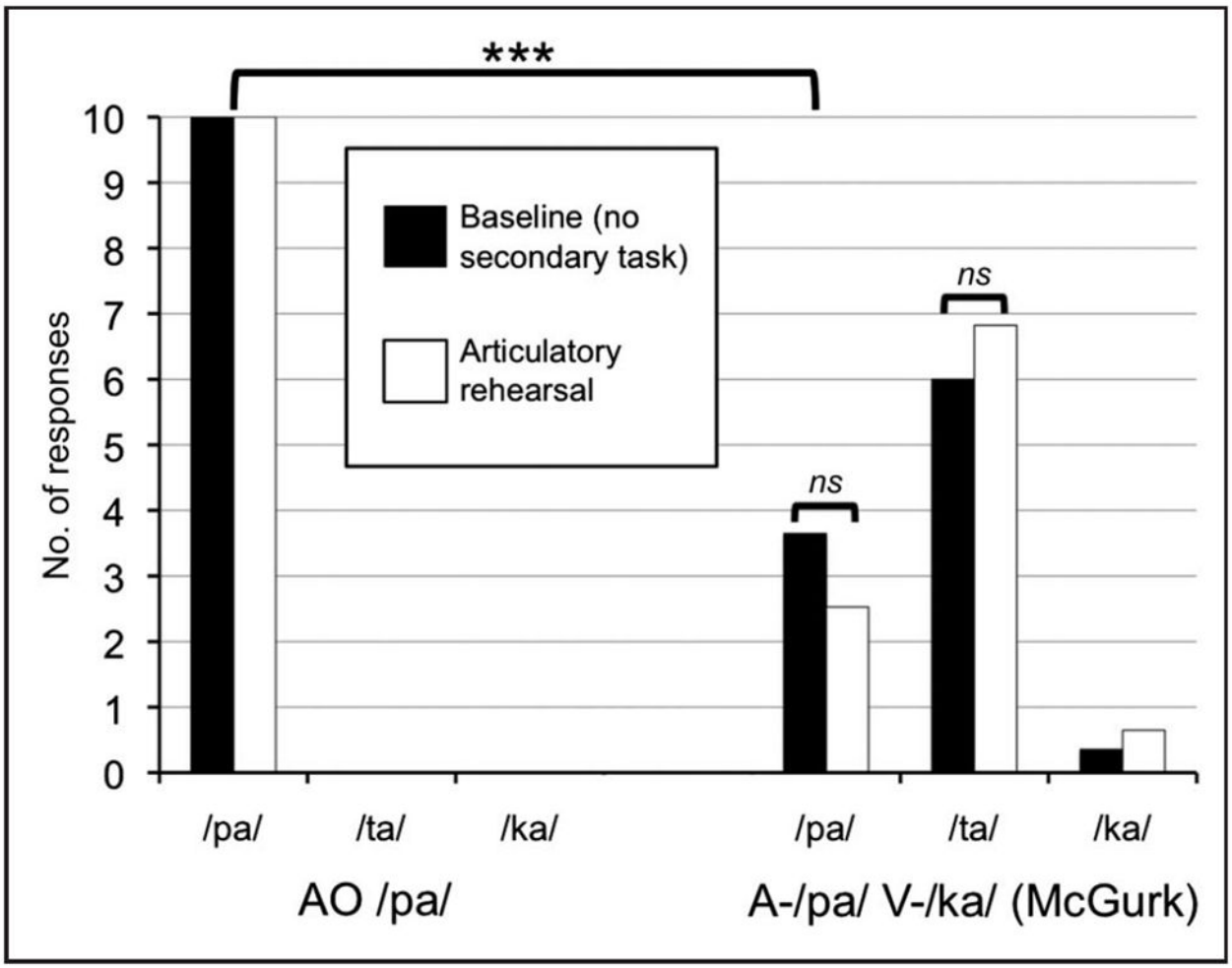

Figure 2.

Average number of responses for each alternative during Experiment 1b for the AO /pa/ and MM (A-/pa/, V-/ka/; McGurk) conditions during baseline (black bars) and the articulatory suppression task (white bars). *ns* = not significant.

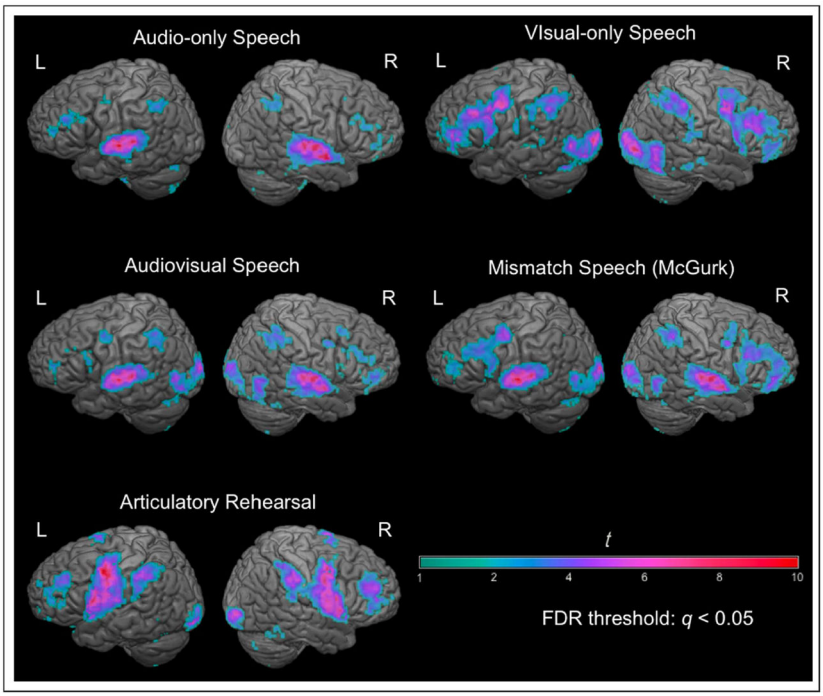

Figure 3.

Activations above baseline during the functional runs from each condition during Experiment 2. All activations are shown with an FDR-corrected threshold of $q < 0.05$. Auditory, AV, and MM speech activated a peri-sylvian language network including superior temporal lobes, inferior/middle frontal gyrus, dorsal precentral gyrus, and inferior parietal lobe. Visual speech activated lateral occipital lobe, posterior middle temporal lobe, inferior/middle frontal gyrus, dorsal precentral gyrus, and inferior parietal lobe. ART activated posterior IFG, precentral gyrus, and inferior parietal lobe.

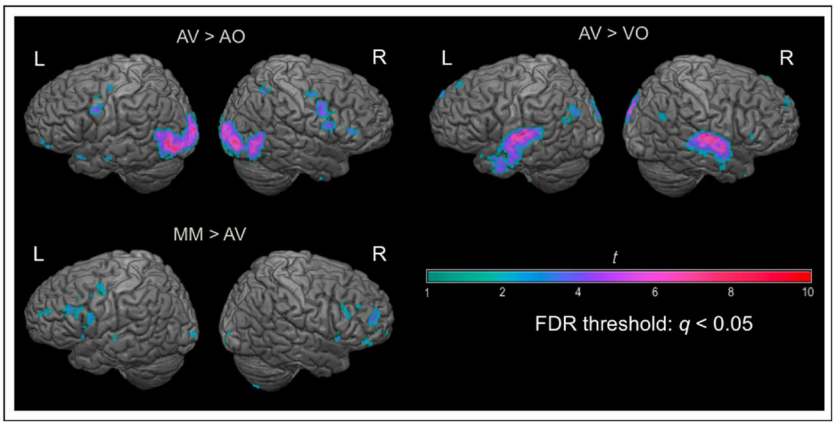

Figure 4.

Contrasts from Experiment 2. All activations are positive with an FDR-corrected threshold of $q < 0.05$.

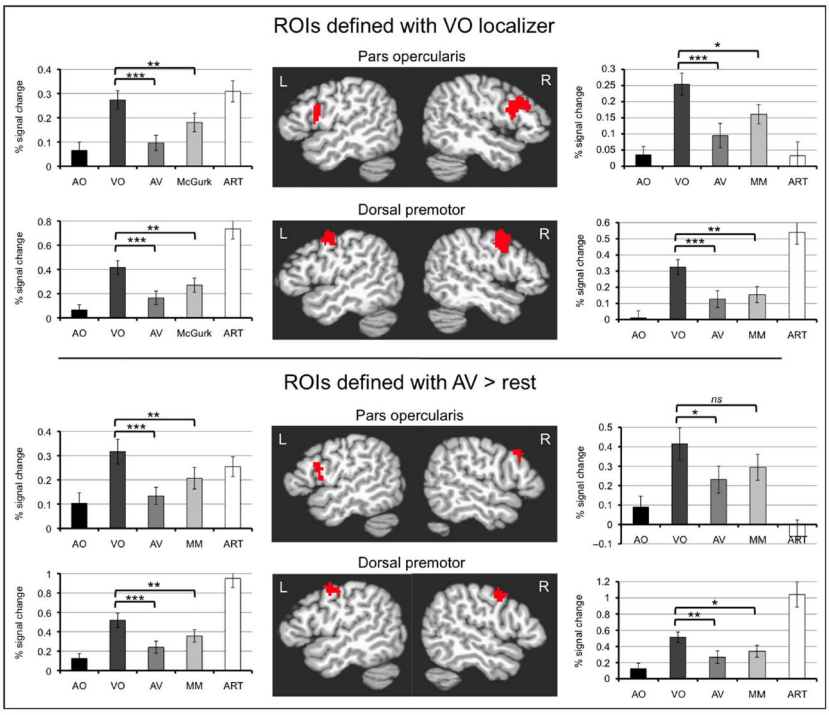

Figure 5.

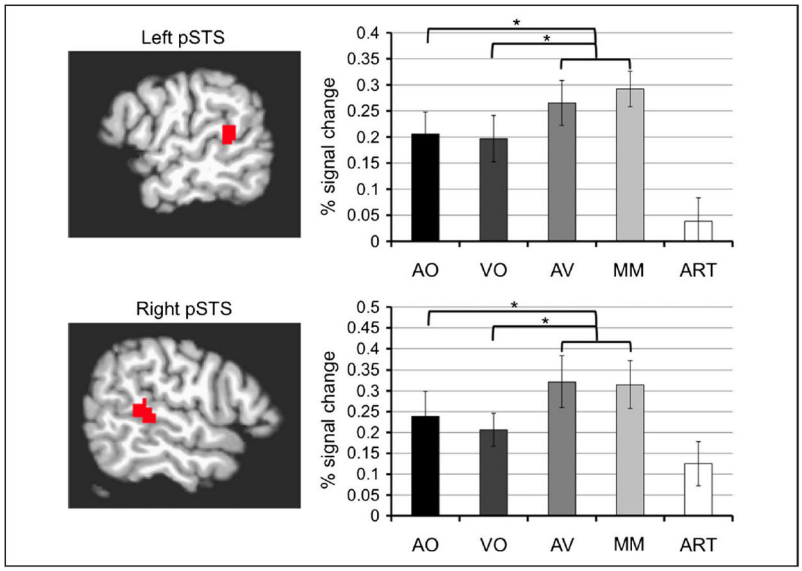

Figure 6.

Analyses of pSTS ROIs from Experiment 2. (top) Left pSTS ROI localized through a conjunction of AO and VO trials during even experimental runs. Percent signal change values from odd runs for each condition within the ROI are reported in the bar graph to the right. The contrast AO: −2 VO: 0 AV: 1 MM: 1 ART: 0 revealed that the multisensory conditions (AV and MM) produced significantly greater activity than AO, F(l, 19) = 6.045, p = .024, and the contrast AO: 0 VO: −2 AV: 1 MM: 1 ART: 0 revealed that the multisensory conditions (AV and MM) produced marginally significantly greater activity than VO, F(1, 19) = 4.748, p = .042. (bottom) Right pSTS ROI localized through the conjunction of AO and VO trials during even experiment runs. Percent signal change values from odd runs for each condition within the ROI are reported in the bar graph to the right. The contrast AO: −2 VO: 0 AV: 1 MM: 1 ART: 0 revealed that the multisensory conditions (AV and MM) produced significantly greater activity than AO, F(1, 19) = 6.045, p = .025, and the contrast AO: 0 VO: −2 AV: 1 MM: 1 ART: 0 revealed that the multisensory conditions (AV and MM) produced marginally significantly greater activity than VO F(1, 19) = 4.748, p = .030. Error bars indicate standard error for each condition. All reported contrasts are Bonferroni corrected for multiple comparisons with a family-wise error rate of $p < .05$ (per-comparison error rates of $p < .025$).

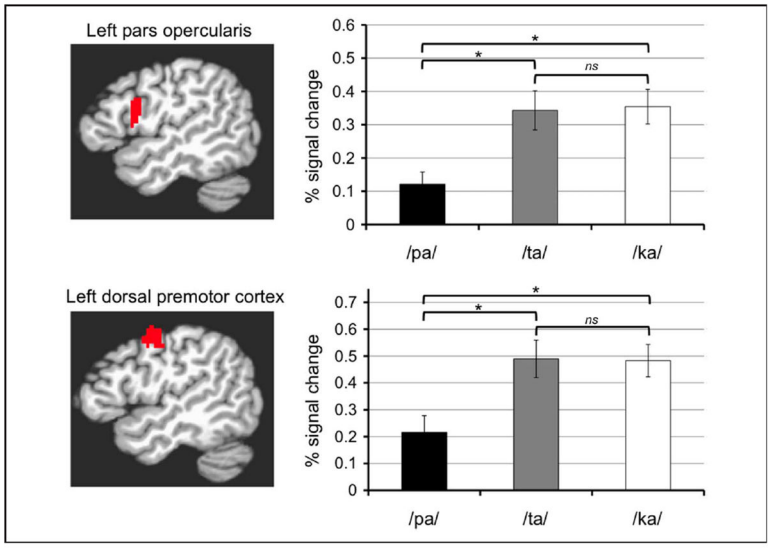

Figure 7.

Analyses of left hemisphere frontal-motor ROIs for the individual stimuli from the VO condition from Experiment 2. (top) Pars opercularis ROI localized during the VO localizer runs. Percent signal change values for each VO stimulus within this ROI are reported in the bar graph to the right. VO /pa/ activated this region significantly less than VO /ta/, $t(19) = -3.392$, $p = .003$, two-tailed, as well as VO /ka/, $t(19) = -4.029$, $p = .001$, two-tailed. There was no significant difference between VO /ta/ and VO /ka/, $t(19) = -0.326$, $p = .748$, two-tailed. (bottom) Dorsal premotor cortex ROI localized during the VO localizer runs. Percent signal change values for each VO stimulus within this ROI are reported in the bar graph to the right. VO /pa/ activated this region significantly less than VO /ta/, $t(19) = -3.634$, $p = .002$, two-tailed, as well as VO /ka/, $t(19) = -3.623$, $p = .002$, two-tailed. There was no significant difference between VO /ta/ and VO /ka/, $t(19) = 0.173$, $p = .864$, two-tailed. Error bars indicate standard error for each condition. *ns* = not significant. All reported *t* tests are Scheffé corrected for post hoc multiple comparisons with a family-wise error rate of $p < .05$ (per-comparison error rates of $p < .008$).

Table 1

Talairach Coordinates of Clusters Activated by the Contrast AV > AO

Region	Hemisphere	x	y	z	Cluster Size (mm^3)
Lateral occipital lobe	Left	−34	−78	−1	22,422
Lateral occipital lobe	Right	37	−74	−1	21,594
Precentral gyrus	Right	51	1	36	1,500
Superior parietal lobule	Right	31	−55	46	1,000
Inferior frontal gyrus (pars opercularis)	Right	55	6	21	890
Inferior frontal gyrus (pars triangularis)	Right	51	29	13	344
Precentral gyrus	Left	−55	−1	35	344
Inferior parietal lobule	Right	38	−36	42	297
Middle OFC	Right	20	43	−9	250

$n = 20$, cluster threshold = 10 voxels, FDR $q < 0.05$.

Table 2

Talairach Coordinates of Clusters Activated by the Contrast AV > VO

Region	Hemisphere	x	y	z	Cluster Size (mm^3)
Medial occipital lobe	Left/right	0	−79	14	22,891
Superior temporal lobe	Left	−51	−20	8	21,781
Superior temporal lobe	Right	56	−17	6	19,422
Angular gyrus	Left	−42	−64	26	2,156
Hippocampus	Left/right	5	−33	−5	1,891
Anterior medial frontal cortex	Left/right	−3	54	7	1,219
Medial superior frontal gyrus	Right	6	55	30	640
Anterior cingulate	Left/right	0	−30	5	313
Angular gyrus	Right	46	−61	25	297
Cerebellum	Left	−19	−48	−19	266
Anterior inferior cingulate	Left/right	1	12	−3	234
Superior medial gyrus	Right	6	46	35	234
BG	Right	25	3	0	188
Orbital gyrus	Right	13	41	6	172
Cerebellum	Left	−19	−50	−46	156

$n = 20$, cluster threshold = 10 voxels, FDR $q < 0.05$.

Table 3

Talairach Coordinates of Clusters Activated by the Contrast MM > AV

Region	Hemisphere	x	y	z	Cluster Size (mm^3)
SMA	Left/right	0	−12	46	4,688
Anterior insula	Left	−32	19	5	3,547
Anterior insula	Right	34	19	3	2,703
Middle frontal gyrus	Right	32	49	13	2,188
Middle/IFG (pars triangularis)	Left	−46	21	27	1,766
Middle frontal gyrus	Left	−31	46	20	1,188
Inferior frontal gyrus	Left	−54	−6	20	672
Precentral gyrus	Left	−47	−5	50	422
Middle/IFG (pars triangularis)	Right	46	20	28	188
Middle/IFG (pars triangularis)	Right	54	24	26	172
Inferior frontal/precentral gyrus	Left	−35	4	30	156

$n = 20$, cluster threshold = 10 voxels, FDR $q < 0.05$).