CrossMark

BRIEF REPORT

No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception

Jessica S. Arsenault[1,2] · Bradley R. Buchsbaum[1,2]

Abstract The motor theory of speech perception has experienced a recent revival due to a number of studies implicating the motor system during speech perception. In a key study, Pulvermüller et al. (2006) showed that premotor/motor cortex differentially responds to the passive auditory perception of lip and tongue speech sounds. However, no study has yet attempted to replicate this important finding from nearly a decade ago. The objective of the current study was to replicate the principal finding of Pulvermüller et al. (2006) and generalize it to a larger set of speech tokens while applying a more powerful statistical approach using multivariate pattern analysis (MVPA). Participants performed an articulatory localizer as well as a speech perception task where they passively listened to a set of eight syllables while undergoing fMRI. Both univariate and multivariate analyses failed to find evidence for somatotopic coding in motor or premotor cortex during speech perception. Positive evidence for the null hypothesis was further confirmed by Bayesian analyses. Results consistently show that while the lip and tongue areas of the motor cortex are sensitive to movements of the articulators,

✉ Bradley R. Buchsbaum
bbuchsbaum@research.baycrest.org

[1] Rotman Research Institute at Baycrest Health Sciences, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1

[2] Department of Psychology, University of Toronto, 100 St. George Street, 4th floor, Toronto, Ontario, Canada M5S 3G3

they do not appear to preferentially respond to labial and alveolar speech sounds during passive speech perception.

There has been a recent resurgence of the motor theory of speech perception (MTSP), the idea stemming from the work of Liberman and colleagues (Liberman, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), arguing that the way we understand speech is by mapping the sounds of speech to the movements associated with producing those sounds (Galantucci, Fowler, & Turvey, 2006).

One study substantially contributing to the revival of this idea is a functional magnetic resonance imaging (fMRI) study by Pulvermüller et al. (2006), which showed that the lip and tongue areas of the premotor/motor cortex (PMC) differentially responded to the perception of lip and tongue sounds during passive listening. This paper has received over 368 citations (Google Scholar) since its publication nearly a decade ago, indicating the considerable influence it has had on the field. While several studies have since demonstrated an effect of transcranial magnetic stimulation (TMS) applied to the lip and tongue areas of PMC on discrimination tasks during speech perception (e.g., D'Ausilio et al., 2009; Möttönen & Watkins, 2009), no fMRI study that we are aware of has yet replicated the specific automatic motor-somatotopic effect reported by Pulvermüller and colleagues.

The objective of the current study was to replicate and extend the findings of Pulvermüller et al. (2006), which showed that the lip and tongue areas of the PMC respond differentially to the passive perception of the phonological feature known as place of articulation (PoA; the location at

which the articulators approach one another to form a speech sound, such as the lips to form a labial consonant).

In the original study, the lip and tongue areas of the motor cortex were first "localized" in 12 participants via a motor task (lip and tongue movements) and an articulation task (silently mouthing the syllables [pI], [pæ], [tI], and [tæ]). The peak voxels in the central sulcus, which were most strongly activated by the localizers, were selected as the centers of 8-mm radius regions of interest (ROIs) for each area, and additional precentral ROIs were also selected. In the perception portion of the study, participants listened to the previously listed speech sounds presented by a single female voice. They were instructed to passively listen to the sounds during two 24-minute blocks without engaging in any motor movements. Spectrally adjusted signal-correlated noise stimuli were also presented to participants.

Mean fMRI activation estimates for the noise stimuli were subtracted from each syllable type and then subjected to a 2 (place of articulation: labial/lip and alveolar/tongue) × 2 (motor cortex location: lip and tongue) analysis of variance (ANOVA) in both the precentral and central ROIs. Results showed a significant interaction of PoA and motor cortex location only in the precentral ROIs, $F(1, 11) = 7.25$, $p = .021$. The authors concluded that articulatory phonetic features are accessed during both passive speech perception and production within the precentral gyrus. They further concluded that the findings were consistent with aspects of the MTSP and that "the different perceptual patterns corresponding to the same phoneme or phonemic feature are mapped onto the same gestural and motor representation" (Pulvermüller et al., 2006, p. 7868).

In a recent study (Arsenault & Buchsbaum, 2015), we used multivoxel pattern analysis (MVPA) to identify areas of the cortex whose distributed activity patterns were associated with phonological features including voicing, manner, and PoA. Although we found robust phonological feature decoding in the bilateral superior temporal area, we did not find significant decoding in PMC. However, in our previous study we did not use an articulatory localizer to identify lip and tongue regions of the PMC and therefore could not claim to directly replicate or refute Pulvermüller et al.'s (2006) results.

The present study used an articulatory localizer and a passive speech perception task to conceptually replicate the findings of Pulvermüller et al. (2006). In addition to the ROI-based univariate approach adopted by the original authors, we also use more sensitive MVPA to identify shared distributed patterns of neural activation between speech production and speech perception in the PMC. If articulatory features are indeed mapped to PMC during passive listening of speech sounds, MVPA ought to be more sensitive to these shared representations than traditional univariate methods.

Materials and methods

Experimental methods

Participants

Thirteen healthy young adults (mean age = 27.5 years, SD = 3.9 years; six females) were recruited from the Baycrest Health Sciences research database. All were right-handed fluent English speakers with normal hearing, no known neurological or psychiatric issues, and no history of hearing or speech disorders. All participants gave written informed consent according to guidelines established by Baycrest's Research Ethics Board.

Stimuli and experimental methods

Stimuli consisted of the eight syllables listed in Table 1 and were perfectly balanced across the phonological features of PoA, manner of articulation, and voicing.

Each participant completed two separate scanning sessions within the span of approximately 1 month (but at least 48 hours apart). The first scanning session consisted of a speech perception task, while the second session consisted of a silent speech production (overt articulation without vocalization; speech miming) task. A third session was also conducted, but this data will not be presented herein. Speech perception was always first so as not to influence the perception of sounds with the associated motor movements (the same protocol was used in Pulvermüller et al., 2006). Before each session, all participants completed a brief practice session outside of the scanner. Each experimental session consisted of either 9 or 10 runs (six and seven participants, respectively) of 27 trials each, with each syllable plus a blank trial appearing three times in each run. Each run was approximately 4.5 minutes in length. Each action trial (either perception or miming, depending on the session) consisted of five stimulus repetitions separated by a 500 ms interstimulus interval (ISI; see Fig. 1). Action trials were interspersed with blank trials, which lasted for 4 s and were visually represented by three dashes (—). All stimuli were presented using Eprime 2.0 and made visible to participants through a mirror mounted on the head coil.

Speech perception session

Auditory recordings were taken from the standardized University of California, Los Angeles, version of the Nonsense-Syllable Test (Dubno & Schaefer, 1992). Each syllable was produced by a female voice (two different speakers; three exemplars of each syllable per speaker) and was root mean-square normalized and adjusted to be 620 ms in length, with a 10 ms rise and fall at the beginning and end of the sound (presented within a 1,000 ms window to allow adequate

Table 1 Phonological feature distribution of stimuli used in current study

Syllable	Place	Manner	Voicing
ba	labial	stop	voiced
pa	labial	stop	voiceless
va	labial	fricative	voiced
fa	labial	fricative	voiceless
da	alveolar	stop	voiced
ta	alveolar	stop	voiceless
za	alveolar	fricative	voiced
sa	alveolar	fricative	voiceless

loading time). The sounds were played at a comfortable volume for participants through electrodynamic MR-Confon headphones. During this session, participants were asked to keep their eyes open while passively listening to speech sounds. To ensure that participants were paying attention, they were asked to make a button press during every blank trial. Responses indicate that participants were attentive throughout the experiment (see Supplementary Material).

Speech miming session

Trials in this silent session included an initial visual presentation of the syllable to be mimed in large black font in the center of the screen, followed by five crosshairs separated by an ISI. In time with each of the crosshairs, participants were asked to mime the syllable appearing in the initial slide as if they were saying the syllable but without actually producing any sound. Thus, they mouthed the syllable five consecutive times during each action trial and rested during blank trials. Participants were trained to keep their heads as still as possible and not to exaggerate the mouthing movements.

Imaging methods

MRI set-up and data acquisition

Participants were scanned with a 3.0-T Siemens MAGNETOM Trio MRI scanner using a 12-channel head coil system. High-resolution gradient-echo multislice T1-weighted scans (160 slices of 1 mm thickness, 19.2 × 25.6 cm field of view) coplanar with the echo-planar imaging scans (EPIs) as well as whole-brain magnetization prepared rapid gradient echo (MP-RAGE) 3-D T1 weighted scans were acquired for anatomical localization, followed by T2*-weighted EPIs sensitive to BOLD contrast. Images were acquired using a two-shot gradient-echo EPI sequence (22.5 × 22.5 cm field of view with a 96 × 96 matrix size, resulting in an in-plane resolution of 2.35 × 2.35 mm for each of 26 3.5-mm axial slices with a 0.5-mm interslice gap; repetition time = 1.5 s; echo time = 27 ms; flip angle = 62 degrees).

MRI preprocessing

Functional images were converted into NIFTI-1 format, realigned to the mean image of the first scan with AFNI's (Cox, 1996) *3dvolreg* program, corrected for slice-timing offsets and spatially smoothed with a 4-mm (full-width half-maximum) Gaussian kernel. To facilitate comparison of fMRI data acquired in the speech perception and speech miming sessions, the EPI images from the second session (miming) were aligned to the reference EPI from the first session (speech perception).

Structural MRIs were warped to MNI space using symmetric diffeometric normalization as implemented using Advanced Normalization Tools (ANTS; Avants, Epstein, Grossman, & Gee, 2008). Each subject's structural MRI was also parcellated into ROIs using FreeSurfer's (Dale, Fischl, & Sereno, 1999) automated automatic anatomical labeling
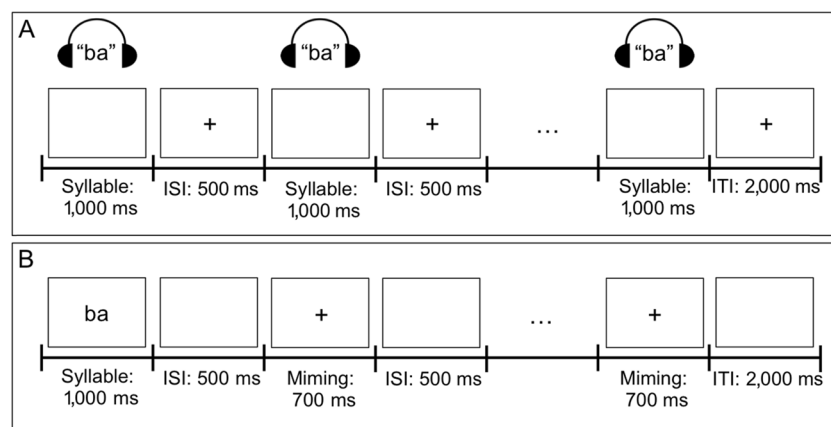

Fig. 1 One action trial each for **a**) speech perception and **b**) speech miming. Participants were asked to simply listen to the sounds during speech perception and to silently mouth the syllables during each of the five crosshairs for speech miming

method ("aparc 2009" atlas; Destrieux, Fischl, Dale, & Halgren, 2010).

Data analysis

Univariate analyses

We first attempted to narrowly replicate the findings from Pulvermüller et al. (2006) by using their exact precentral ROI MNI coordinates (lips: -54, -3, 46; tongue: -60, 2, 25) surrounded by a 10-mm spherical mask. We estimated BOLD activity for each syllable using the general linear model (GLM) with a set of eight regressors of interest, one per syllable, derived by convolving the SPM canonical hemodynamic response function with the event onset vector (e.g., the times at which each syllable was mimed or perceived). An additional set of five orthogonal polynomial nuisance regressors were entered into the model and used to model low-frequency drift in the fMRI time series.

Using the GLM regression weights, averaged over the spherical mask, for two of our eight stimuli – [pæ] and [tæ] – we performed a 2 (stimulus type: labial and alveolar) × 2 (motor location: lip and tongue) within-subjects ANOVA. We also investigated PoA sensitivity in these coordinates using all eight stimuli (four labial, four alveolar) with one-way t tests.

We followed this with a general test of PoA sensitivity in PMC during speech perception using subject-specific somatotopic localizers. To isolate the lip and tongue areas of the PMC, we contrasted regression weights for labial and alveolar PoA using all eight syllables during the speech miming task and yielding voxelwise t statistics. For each participant, we identified the peak positive (labial > alveolar contrast) and negative (alveolar > labial contrast) voxel in both the left and right PMC (see Fig. 2a and Supplementary Material Table S1 for individual coordinates). We then created 10-mm radius spheres around each of the peak voxels to use as lip and tongue masks. Because of the close physical proximity of these locations, we ensured that the lip area was not included in the tongue mask and vice versa by removing any voxels in each participants' mask that contained negative (alveolar > labial) values for the lip ROIs and positive (labial > alveolar) values for the tongue ROIs. We used these lip and tongue ROIs to extract regression weights from the speech perception data to assess the extent to which the passive perception of labial and alveolar sounds is associated with elevated activation in the lip and tongue areas, respectively. We conducted paired t tests of labial versus alveolar sounds within each ROI.

Multivariate analyses

MVPA identifies patterns of activation that are distributed across a region and can be used to reliably classify brain patterns according to experimental condition, cognitive state, or other variables (Norman, Polyn, Detre, & Haxby, 2006). Here we used MVPA to test whether the auditory perception of speech sounds selectively activates the distributed motor speech representations involved in the production of those same sounds. This is a more general test of the conclusions reached by Pulvermüller and colleagues (2006), who stated that the "perception of speech sounds in a listening task activates specific motor circuits in precentral cortex, and [these] are the same motor circuits that also contribute to articulatory and motor processes invoked when the muscles generating the speech sounds are being moved" (p. 7868). Thus, we used the articulation task as a method of training a pattern classifier that could be used to test the reliability of PoA encoding in the PMC during silent articulation and, more importantly, during passive perception.

To achieve this, we first trained a multivariate classifier on the miming data (sparse partial least squares implemented in the R package "spls": https://cran.r-project.org/web/packages/spls; Chun & Keles, 2007) in each of six FreeSurfer subject-specific anatomical PMC ROIs (the left and right precentral gyrus, central sulcus, and postcentral gyrus; see Fig. 2b) to discriminate between alveolar and labial tokens in the full set of eight syllables. To carry out an MVPA analysis, one must first estimate activation magnitudes for each trial in the experiment. Thus, we first ran new GLM analyses for both the perception and articulation data sets in which each trial was modeled with an individual regressor that was time-locked to a single experimental event. This analysis yielded a beta coefficient image for every trial in the experiment, resulting in two sets of 216 beta images (216 speech perception; 216 speech miming; note, for subjects with 10 runs, the number of trials were 240 perception and 240 miming. For the sake of simplicity, we will use the more conservative values of nine runs / 216 images to describe the methods). The set of 216 beta images from the miming task was used to train a pattern classifier to discriminate between the labial and alveolar beta images.

To assess the performance of this PoA classifier, we used a cross-validation approach where trial-wise activation images from eight runs (192 trials) were used as a training set and the remaining data (24 trials) were used as a hold-out test set (see Arsenault & Buchsbaum, 2015, for similar approach). This process was then repeated such that the beta images for each run were used as the heldout test set for one iteration of the cross-validation procedure. The results of this internal cross-validation of the classification accuracy on the speech miming data established our ability to reliably discriminate between labial and alveolar articulations during miming of syllables (see Fig. 4). We then fit a final model on all 216 speech miming trials and stored the fitted model for later use (see below).

To test whether the classifier trained on the speech miming beta images could also be used to accurately classify speech *perception* beta images, we adopted a "cross-decoding" approach (Kriegeskorte, 2011). This involves training a pattern classifier on one condition (i.e., speech miming) and testing

the fitted model on another condition (i.e., speech perception). The cross-decoding approach has been used successfully in vision literature, for example, to reveal overlapping patterns of activation between perceptual data and visual imagery (Stokes, Thompson, Cusack, & Duncan, 2009; St-Laurent, Abdi, & Buchsbaum, 2015). If speech miming and speech perception share a representational code in PMC, then the PoA classifier trained on speech *miming* data should be able to reliably classify PoA in the speech *perception* data. Above-chance classification from miming to speech perception would indicate that the classifier generalizes from speech ar-ticulation to speech perception, and thus the two neural states are related by a common pattern of activity. We applied the stored speech miming classification model to the set of 216 beta images derived from the perception data and calculated the classification accuracy of the recorded output. Two-tailed *t* tests were computed to test if the classification accuracy in each hemisphere was significantly different than chance (0.5), and were corrected for multiple comparisons using a false discovery rate (FDR; Benjamini & Hochberg, 1995).

Finally, we trained a second model using the speech per-ception data alone to assess whether reliable patterns of acti-vation associated with PoA exist in PMC that are driven by acoustic input without the constraint that the evoked patterns are topographically related to the patterns evoked during speech miming. To evaluate this possibility, we used the same cross-validation procedure as described above, except that this time the classifier was trained and tested on the speech per-ception beta images instead of being trained on the speech miming beta images. To distinguish this analysis from the cross-decoding approach presented above, we refer to it as a *direct-decoding* analysis. We then examined performance of this classifier within the PMC ROIs using only the auditory-speech data and used two-tailed *t* tests to assess whether per-formance in each ROI was significantly different from chance (0.5).

Bayesian hypothesis testing

Bayesian hypothesis testing was used throughout this report to assess the relative statistical evidence for the null hypothesis compared to the alternative by taking the reciprocal of the Bayes factor (1/Bayes Factor; Rouder, Speckman, Sun, Morey, & Iverson, 2009). Bayes Factors represent the relative weight of evidence in the data favoring one hypothesis over another given appropriate prior distributions. Here we used the implementation of Bayesian *t* tests in the R package "BayesFactor" (https://cran.r-project.org/web/packages/BayesFactor/). The alternative hypothesis was specified using a Cauchy distribution such that 50 % of the true effects sizes are within the interval (-0.7071, 0.7071). This is a reasonable choice in the present case where we wish to

compare the null hypothesis to an alternative in which the effect size is small to medium (see Rouder et al., 2009).

Results

Univariate results

Previously reported coordinates

A two-way within-subjects ANOVA was performed to test for an interaction of motor areas (lip and tongue) and stimulus type (/p/ and /t/) within the MNI coordinates reported in Pulvermüller et al. (2006). While the effect was significant for the speech miming task, $F(1, 12) = 17.99$, $p < .001$, we did not find a significant interaction for speech perception, $F(1, 12) = 0.57$, $p = .46$. These coordinates also failed to produce a significant difference in an overall test of PoA con-trast (labial > alveolar) using all eight syllables for speech perception in our participants—lip area: $t(12) = 0.10$, $p = .92$, 1/Bayes Factor = 3.58; tongue area: $t(12) = 0.22$, $p = .83$, 1/Bayes Factor = 3.52. Bayesian hypothesis tests suggest that evidence for the null hypothesis (that lip and tongue areas of PMC do not selectively respond to labial and alveolar speech sounds) is approximately 3.5 times larger than for the alternative (that lip and tongue areas do selectively respond to labial and alveolar sounds).

Functionally derived coordinates

We next used our participants' individually localized bilateral lip and tongue ROIs to test for somatotopic activation during passive speech perception. We were able to identify in the articulatory localizer data significant lip and tongue PMC ROIs in all 13 subjects (all single-subject $ps < .0001$; see Fig. 2a for evidence of high spatial overlap). We conducted paired *t* tests to test for a difference in activation between labial and alveolar articulations in both the lip and tongue ROIs. These tests confirmed that labial and alveolar move-ments produced significantly different activation values dur-ing the speech miming task in both the lip area ($M_{labial} = 0.526$, $SD = 0.285$; $M_{alveolar} = 0.299$, $SD = 0.269$); $t(12) = 10.08$, $p < .0001$, 95 % CIs [0.179, 0.277], $d = 0.824$, and the tongue area ($M_{labial} = 0.119$, $SD = 0.329$; $M_{alveolar} = 0.349$, $SD = 0.34$); $t(12) = -9.43$, $p < .0001$, 95 % CIs [-0.283, -0.177], $d = -0.687$, which is to be expected because the ROIs were selected on the basis of a labial > alveolar GLM contrast. As can be seen in Fig. 3, however, no evidence was found for a significant difference between the auditory perception of labial and alveolar sounds in either the lip area ($M_{labial} = 0.035$, $SD = 0.064$; $M_{alveolar} = 0.031$, $SD = 0.05$); $t(12) = 0.251$, $p = .805$, 95 % CIs [-0.027, 0.034], or the tongue area ($M_{labial} = 0.026$, $SD = 0.088$; $M_{alveolar} = 0.027$, $SD = 0.072$);
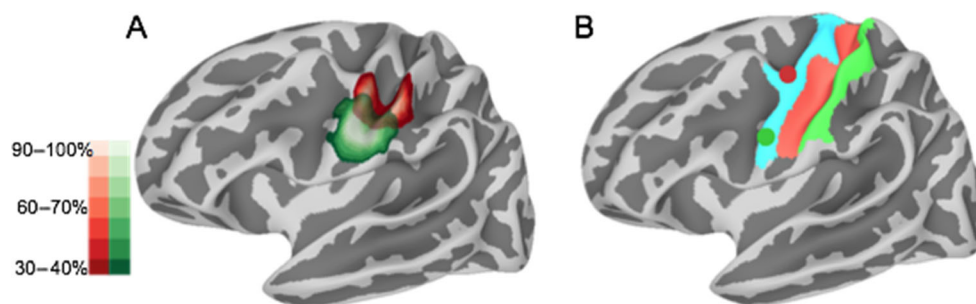
Fig. 2 ROIs used in the current study projected onto inflated brain surface. **a** Average of each participant's peak lip (red) and tongue (green) voxels, plus 10-mm spherical mask, derived from the localizer task. Color bars indicate percentage of overlap between participants. **b** FreeSurfer ROIs corresponding to precentral gyrus (cyan), central sulcus (pink), and postcentral gyrus (lime green). Red and green dots indicate approximate location of Pulvermüller et al.'s (2006) lip and tongue coordinates, respectively. (Color figure online)

$t(12) = -0.112$, $p = .913$, 95 % CIs [-0.026, 0.003]. Furthermore, Cohen's effect size suggests low practical significance during the speech perception task for both the lip ($d = 0.062$) and tongue ($d = -0.016$) ROIs, and the overall interaction of PoA and ROI was not significant, $F(1, 12) = 0.492, p = 0.496$. Comparable effect size data is unfortunately not available from Pulvermüller et al. (2006).

This finding was not affected by hemisphere as no significant differences were found when the left and right ROIs were investigated separately (see Supplementary Figure S1).

Multivariate analyses

We trained a sparse partial least squares (Chun & Keles, 2007) pattern classifier in three PMC ROIs to classify PoA on the speech miming data and tested the performance of this classifier on both the speech miming and the speech perception data. Figure 4 shows classification accuracy for the left and right precentral gyrus, central sulcus, and postcentral gyrus, averaged across participants. The results show that when our speech miming classifier was tested on the held-out speech miming data, it produced highly significant classification accuracy (mean accuracy: 0.86, ranging from 0.81 to 0.92 across

ROIs; chance accuracy: 0.5; all participants' average accuracies > 0.75; all ps < .0001). When this model was used to predict PoA for the speech perception task, however, it failed to reliably classify PoA in any of the ROIs (ps > 0.5; see Table 2 for statistical results). This suggests again that the passive perception of labial and alveolar speech sounds does not recruit the same motor patterns as the corresponding articulatory movements. Our Bayesian hypothesis tests revealed that across ROIs, evidence for the null hypothesis (that classifier accuracy is equal to 0.5) is 2.3 to 3.6 times larger than for
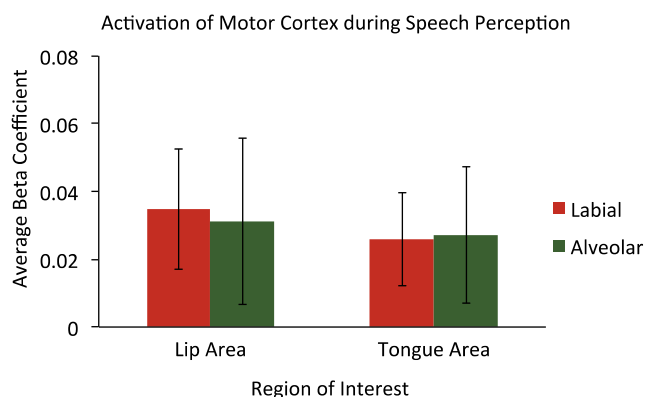

Fig. 3 Univariate analysis of speech perception task, showing average activation values within each participants' lip and tongue area for labial and alveolar speech sounds. Error bars represent between-subjects standard error of the mean
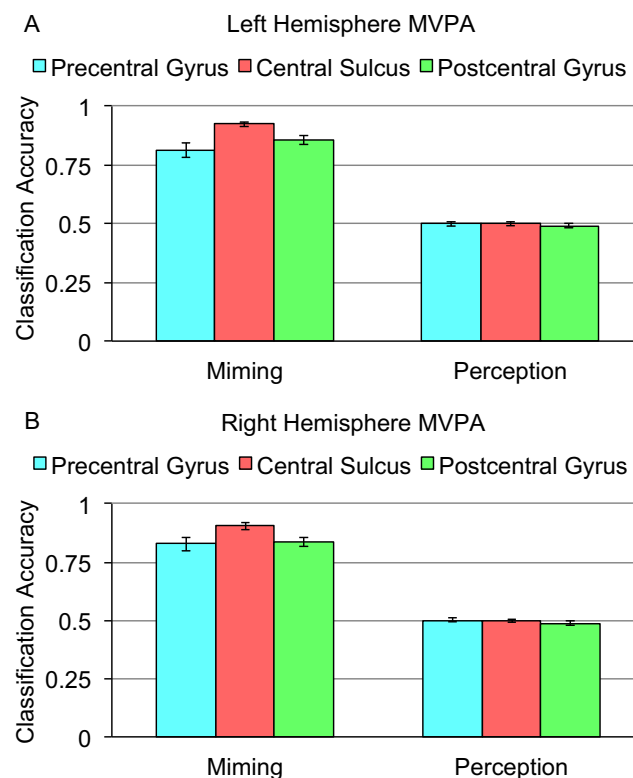

Fig. 4 Classification accuracy for a classifier trained on speech miming data and tested on both speech miming and speech perception (cross-decoding). Chance performance = 0.5. **a** Left hemisphere FreeSurfer ROIs. **b** Right hemisphere FreeSurfer ROIs

Table 2 MVPA cross-validation, cross-decoding, and Bayes factor results. Bayes factor results represent how much more likely the null hypothesis is compared to the alternative

ROI	Miming Mean	Miming t value (df = 12)	Perception Mean	Perception t value (df = 12)	1/Bayes Factor
Left precentral gyrus	0.8123	10.5995*	0.5003	0.0291 (ns)	3.59
Left central sulcus	0.9238	34.0952*	0.4997	-0.0369 (ns)	3.59
Left postcentral gyrus	0.8569	18.5625*	0.4907	-1.0115 (ns)	2.33
Right precentral gyrus	0.8284	11.0143*	0.5014	0.1961 (ns)	3.53
Right central sulcus	0.9054	25.8642*	0.4982	-0.2387 (ns)	3.51
Right postcentral gyrus	0.8362	16.8201*	0.4882	-0.9795 (ns)	2.39

* < .0001; all p values adjusted for multiple comparisons with false discovery rate correction

the alternative hypothesis (that classifier performance is greater than 0.5).

Direct decoding

In a final analysis, we investigated cross-validated classification accuracy for just the speech perception data to assess whether there were neural patterns in the PMC that could discriminate the speech tokens without the requirement that these patterns be congruent with the corresponding speech miming patterns (and thus would not have been detected during cross-decoding). Classification accuracy for this analysis was not significantly different than chance in any region after adjusting for multiple comparisons, although there is marginal effect in the left postcentral gyrus at an uncorrected threshold (mean across ROIs = 0.504, chance = 0.5, ps > 0.1; see Fig. 5 and Table 3).

Discussion

The objective of the current study was to replicate and extend the results of Pulvermüller and colleagues (2006), which showed that the passive auditory perception of labial and alveolar consonants differentially activated areas of PMC that

Direct-Decoding Classification Accuracy

Fig. 5 Direct-decoding performance in motor cortex ROIs for a classifier that was trained and tested on speech perception data

are involved in their silent production. We attempted to generalize the finding to a larger set of speech tokens and also to apply a more sensitive MVPA analytic approach searching for somatotopic representations in PMC during passive speech perception. Using a silent articulation localizer, we were able to correctly classify brain patterns evoked during the silent production of labial and alveolar sounds at up to 92 % accuracy at the single trial level. However, both univariate and multivariate analyses failed to find evidence for articulatory representations during passive speech perception in either previously reported, functionally derived, or anatomically defined ROIs within the PMC.

Although it does not appear that PMC selectively maps PoA during speech perception, a large body of research suggests that other areas of the brain – particularly the superior temporal lobes – are sensitive to phonetic properties and specifically encode phonological features such as PoA, voicing, and manner of articulation (Agnew, McGettigan, & Scott, 2011). Recent research by Arsenault and Buchsbaum (2015) used fMRI and MVPA to show overlapping but distinct representations of each feature throughout the auditory cortex, indicating that pattern classification of such perceptual features is indeed possible. These findings were recently largely replicated by Correia, Jansma, and Bonte (2015; see below). Lawyer and Corina (2014) have used fMRI-adaptation to show that voicing and PoA are processed in separate but adjacent areas of the superior temporal lobes in a large stimulus set with multiple consonant and vowel combinations. Similarly, using electrocorticography (ECoG), Mesgarani, Cheung, Johnson, and Chang (2014) showed that electrodes throughout the superior temporal gyrus (STG) selectively respond to particular features during continuous speech, although the nature of ECoG did not allow for the PMC to be assessed in that study.

On the other hand, there have been a number of TMS studies that have suggested that the PMC does have some influence on the perception of speech sounds. D'Ausilio and colleagues (2009) administered TMS to lip- and tongue-relevant motor areas during the discrimination of lip- and tongue-articulated speech sounds and found a double
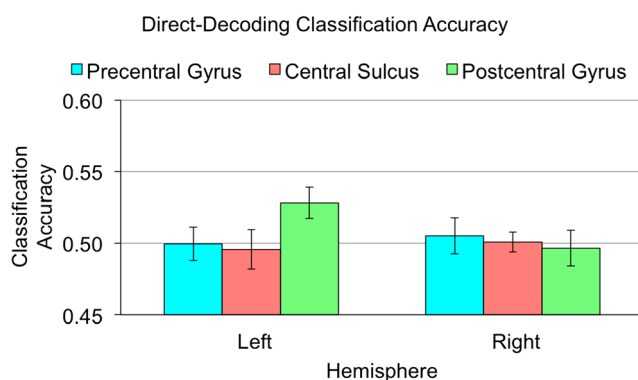
Table 3 Means (chance = 0.5) and *t* values of direct-decoding classification accuracy in PMC ROIs during speech perception; adjusted for multiple comparisons using false discovery rate correction. Bayes factor scores represent evidence in favor of the null hypothesis

ROI	Mean	*t* value (*df* = 12)	Corrected *p* value	1/Bayes Factor
Left precentral gyrus	0.4997	-0.025	0.98	3.59
Left central sulcus	0.4958	-0.2997	0.98	3.46
Left postcentral gyrus	0.5283	2.5849	0.143	0.35
Right precentral gyrus	0.5052	0.4103	0.98	3.34
Right central sulcus	0.5009	0.1319	0.98	3.57
Right postcentral gyrus	0.4967	-0.2593	0.98	3.49

dissociation in reaction time performance. More recently, Schomers, Kirilina, Weigand, Bajbouj, and Pulvermüller (2014) delivered facilitory TMS to lip and tongue areas of the motor cortex during a word-to-picture matching task in order to measure comprehension effects. It was found that words with an alveolar PoA were matched faster following TMS to the tongue area of the motor cortex than to the lip area, although the reverse was not found for labial PoA, and no differences in accuracy were observed.

How can we account for these seemingly contradictory findings? One important caveat to consider when interpreting PMC contribution is the nature of the task (Scott, McGettigan, & Eisner, 2009). For example, Burton, Small, and Blumstein (2000) found that inferior frontal gyrus (IFG) activation was associated with same–different judgements of spoken words only when segmentation – separating an individual sound from the whole stimulus – was required to complete the task. Similarly, a TMS study by Krieger-Redwood, Gaskell, Lindsay, and Jefferies (2013) attempted to disentangle decision-making tasks by using real-word stimuli that differed in only one phoneme (i.e., *carp* vs. *cart*). They asked participants to make either a semantic or phonological judgement and concluded that PMC is involved in phonological, but not semantic, judgements, demonstrating a lack of participation of PMC in speech comprehension. More recently, Möttönen, Van de Ven, and Watkins (2014) used TMS and MEG to measure the impact of attention on auditory-motor processing. They found that in the absence of a task, TMS to motor cortex resulted in nonspecific modulations of auditory cortex; articulator-specific effects of TMS were only seen during a working memory task. This line of research suggests that the above TMS studies, which did not have control tasks, can be explained by their use of phonological categorization (but see Schomers et al., 2014, for their interpretation of the Krieger-Redwood study). It is currently unclear how fMRI-adaptation approaches, which have shown greater effects of phonemic category (e.g., Chevillet et al., 2013), can be reconciled with MVPA approaches such as ours, which show little or no sensitivity to phonological categories in motor cortex.

Recent lesion work also supports the notion that the motor cortex is not critical for passive perception but rather supports

explicit categorization. Stasenko and colleagues (2015) presented evidence that a stroke patient with damage to his left IFG has intact perceptual discrimination but impaired phonemic identification, once again suggesting that the role of the motor system in speech perception has more to do with task demands than speech processing.

The original Pulvermüller et al. (2006) study did not, however, require any judgement on the part of the participant – the task was simply to passively listen, as was done in the current study. Obviously, this type of design would not be practical for a TMS study, but it is noteworthy that much of the recent support for the MTSP comes from TMS, rather than fMRI, studies. Many fMRI studies support the idea that the speech motor system may aid in speech perception under adverse listening conditions (i.e., Osnes, Hugdahl, & Specht, 2011; Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012). Du et al. (2014) demonstrated PMC recruitment during a syllable-identification task that exceeded the contribution from the STG only when speech was presented in white noise. This highlights a difference between our study and the Pulvermüller et al. (2006) study, in that Pulvermüller et al. used a sparse-sampling fMRI paradigm, reducing the level of background scanner noise during the study, while we did not. According to previous literature, the background scanner noise in our study should actually have *increased* the role of the PMC in speech perception, yet we still failed to find a significant somatotopic effect in our task.

Another important aspect of the original Pulvermüller et al. (2006) study is the fact that even they did not find overlap between lip and tongue areas identified by the localizer tasks and the perception of labial and alveolar sounds. Rather, the only significant findings reported in their study were in the precentral ROIs, which were placed anteriorly to the ROIs defined by the localizers, but still within the ROIs we examined (see Fig. 2).

One recent fMRI study used MVPA to show decoding of place of articulation in somatosensory cortex of the inferior postcentral gyrus during passive listening (Correia et al., 2015). It is important to note that this effect was observed using a "direct decoding" approach using acoustic speech stimuli, similar to our second MVPA analysis presented in

Fig. 5. Indeed, our finding of marginally significant direct decoding of PoA in left postcentral gyrus during speech perception may be viewed as a partial replication of the finding in Correia et al. (2015). We did not, however, find that the postcentral gyrus or the other PMC ROIs contained neural pattern information that could be cross-decoded from the speech miming condition, which is the critical test of motor theories of speech perception.

In the current study, we tested the hypothesis that PMC maps articulatory features of speech sounds in areas that are associated with the production of such sounds, as well as within anatomically defined ROIs and in the coordinates specified by Pulvermüller et al. (2006). We did not find any evidence of articulatory feature mapping for PoA in the PMC during passive speech perception and therefore failed to confirm the earlier finding.

References

Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and non-speech mouth sounds. *Journal of Cognitive Neuroscience, 23*(12), 4038–4047. doi:10.1162/jocn_a_00106

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *Journal of Neuroscience, 35*(2), 634–642. doi:10.1523/JNEUROSCI.2454-14.2015

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis, 12*(1), 26–41.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological), 57*(1), 289–300.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12*(4), 679–690. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10936919

Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience, 33*(12), 5208–5215. doi:10.1523/JNEUROSCI.1870-12.2013

Chun, H., & Keles, S. (2007). Sparse partial least squares for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society, Series B, 72*(1), 3–25.

Correia, J. M., Jansma, B., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *The Journal of Neuroscience, 35*(45), 15015–15025.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*(3), 162–173. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8812068

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology: CB, 19*(5), 381–385. doi:10.1016/j.cub.2009.01.017

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis I. Segmentation and surface reconstruction. *NeuroImage, 194,* 179–194.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage, 53*(1), 1–15. doi:10.1016/j.neuroimage.2010.06.010

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America, 111*(19), 7126–7131. doi:10.1073/pnas.1318738111

Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *Journal of the Acoustical Society of America, 91*(4 Pt. 1), 2110–2121. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1597602

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361–377.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language & Cognitive Processes, 27*(7/8), 1145–1166. doi:10.1080/01690965.2012.662280

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., Jefferies, E. (2013). The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, 2179–2188. doi:10.1162/jocn

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage, 56*(2), 411–421. doi:10.1016/j.neuroimage.2011.01.061

Lawyer, L., & Corina, D. (2014). An investigation of place and voice features using fMRI-adaptation. *Journal of Neurolinguistics, 27*(1), 18–30. doi:10.1016/j.jneuroling.2013.07.001

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America, 29*(1), 117–123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*(6174), 1006–1010. doi:10.1126/science.1245994

Möttönen, R., Van de Ven, G. M., & Watkins, K. E. (2014). Attention fine-tunes auditory–motor processing of speech. *The Journal of Neuroscience, 34*(11), 4064–4069. doi:10.1523/JNEUROSCI.2214-13.2014

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience, 29*(31), 9819–9825. doi:10.1523/JNEUROSCI.6018-08.2009

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424–430.

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage, 54*(3), 2437–2445. doi:10.1016/j.neuroimage.2010.09.078

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(20), 7865–7870. doi:10.1073/pnas.0509989103

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null

hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. doi: 10.3758/PBR.16.2.225

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., Pulvermüller, F. (2014). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cerebral Cortex*, 1–9. doi:10.1093/cercor/bhu274

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10,* 295–302.

Stasenko, A., Bonn, C., Teghipco, A., Garcea, F. E., Sweet, C., Dombovy, M., … Mahon, B. Z. (2015). A causal test of the motor theory of speech perception: a case of impaired speech production and spared speech perception. *Cognitive Neuropsychology, 32*(2), 38–57. doi: 10.1080/02643294.2015.1035702

St-Laurent, M., Abdi, H., Buchsbaum, B. R. (2015). Distributed patterns of reactivation predict vividness of recollection. *Journal of Cognitive Neuroscience*, 1–10. doi:10.1162/jocn

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience, 29*(2009), 1565–1572. doi:10.1523/JNEUROSCI.4657-08.2009