eNeuro

Research Article: New Research | Cognition and Behavior

Sensorimotor Representation of Speech Perception — Cross-Decoding of Place of Articulation Features during Selective Attention to Syllables in 7T fMRI

Mario E. Archila-Meléndez[1,2,6], Giancarlo Valente[1,2,6], Joao Correia[1,2,6], Rob P. W. Rouhl[3,4,6], Vivianne H. van Kranen-Mastenbroek[5,6] and Bernadette M. Jansma[1,2,6]

[1]*Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, Maastricht, 6229 EV, The Netherlands*

[2]*Maastricht Brain Imaging Center (M-BIC), Maastricht University, Oxfordlaan 55, Maastricht, 6229 EV, The Netherlands*

[3]*Department of Neurology, Maastricht University Medical Center, Maastricht, 6202 AZ, The NetherlandsP.O. Box 5800*

[4]*School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, 6200 MD, The NetherlandsP.O. Box 616*

[5]*Department of Clinical Neurophysiology, Maastricht University Medical Center, P. Debyelaan 25, Maastricht, 6229 HX, The Netherlands*

[6]*Center for Integrative Neuroscience (CIN), Maastricht University, 6200 MD, Maastricht, P.O. Box 616, The Netherlands*

DOI: 10.1523/ENEURO.0252-17.2018

Received: 17 July 2017

Revised: 9 February 2018

Accepted: 14 February 2018

Published: 22 March 2018

Author Contributions: MAM, GV, JC and BJ designed the research; MAM performed the research; GV, JC contributed with analytic tools; MAM, JC and GV analyzed the data; MAM, GV, JC, RR, VKM and BMJ wrote the paper.

Funding: http://doi.org/10.13039/100007637Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS)
568

Funding: University Fund Limburg/SWOL

The authors declare no competing financial interests.

This work was supported by the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS), call number 568, and by University Fund Limburg/SWOL.

Correspondence should addressed to Mario E. Archila-Meléndez, Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands, E-mail: m.archilamelendez@maastrichtuniversity.nl

Cite as: eNeuro 2018; 10.1523/ENEURO.0252-17.2018

Sensorimotor representation of speech perception – Cross-decoding of place of articulation features during selective attention to syllables in 7T fMRI

2. Abbreviated Title: Sensorimotor representation of speech perception.

3. Authors Names and Affiliations: Mario E. Archila-Meléndez[1, 2, 6], Giancarlo Valente[1, 2, 6], Joao Correia[1, 2, 6], Rob P.W. Rouhl[3, 4, 6], Vivianne H. van Kranen-Mastenbroek[5, 6], Bernadette M. Jansma[1, 2, 6].

[1]: Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands.

[2]: Maastricht Brain Imaging Center (M-BIC), Maastricht University, Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands.

[3]: Department of Neurology, Maastricht University Medical Center, P.O. Box 5800, 6202 AZ, Maastricht, The Netherlands.

[4]: School for Mental Health and Neuroscience (MHeNS), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

[5]: Department of Clinical Neurophysiology, Maastricht University Medical Center, P. Debyelaan 25, 6229 HX, Maastricht, The Netherlands.

[6]: Center for Integrative Neuroscience (CIN), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

4. Author Contributions: MAM, GV, JC and BJ designed the research; MAM performed the research; GV, JC contributed with analytic tools; MAM, JC and GV analyzed the data; MAM, GV, JC, RR, VKM and BMJ wrote the paper.

5. Correspondence should addressed to: Mario E. Archila-Meléndez, Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV, Maastricht, The Netherlands, E-mail: m.archilamelendez@maastrichtuniversity.nl

6. Number of Figures: 5

7. Number of Tables: 0

8. Number of Multimedia: 0

9. Number of words for Abstract: 249

10. Number of words for Significant Statement: 107

11. **Number of words for Introduction:** 672

12. Number of words for Discussion: 1648

13. Acknowledgments: We thank Lars Hausfeld for technical assistance.

14. Conflict of interest: The authors declare no competing financial interests.

15. Funding sources: This work was supported by the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS), call number 568, and by University Fund Limburg/SWOL.

Abstract

Sensorimotor integration, the translation between acoustic signals and motoric programs, may constitute a crucial mechanism for speech. During speech perception, the acoustic-motoric translations include the recruitment of cortical areas for the representation of speech articulatory features, such as place of articulation. Selective attention can shape the processing and performance of speech perception tasks. Whether and where sensorimotor integration takes place during attentive speech perception remains to be explored. Here, we investigate articulatory feature representations of spoken consonant-vowel syllables during two distinct tasks. Fourteen healthy humans attended to either the vowel or the consonant within a syllable in separate delayed-match-to-sample tasks. Single-trial fMRI BOLD responses from perception periods were analyzed using multivariate pattern classification and a searchlight approach to reveal neural activation patterns sensitive to the processing of place of articulation (i.e., bilabial/labiodental vs. alveolar). To isolate place of articulation representation from acoustic covariation, we applied a cross-decoding (generalization) procedure across distinct features of manner of articulation (i.e., stop, fricative, and nasal). We found evidence for the representation of place of articulation across tasks and in both tasks separately: for attention to vowels, generalization maps included bilateral clusters of superior and posterior temporal, insular, and frontal regions; for attention to consonants, generalization maps encompassed clusters in temporoparietal, insular, and frontal regions within the right hemisphere only. Our results specify the cortical representation of place of articulation features generalized across manner of articulation during attentive syllable perception, thus supporting sensorimotor integration during attentive speech perception and demonstrating the value of generalization.

Significance Statement

Speech is supported by sensorimotor integration, a bidirectional translation of its auditory and motoric signals. Whether our brain represents speech as articulatory features during selective attention has not yet been well specified. We focused on the representation of articulatory information of speech during attentive speech perception. For the first time, we applied generalization in classification analysis to counteract the differences in acoustic properties that accompany articulatory information. Participants attended to either the vowels or consonants of syllables, while undergoing fMRI. Our results show that articulatory information is represented in widespread cortical areas during selective attention to the different syllable components, supporting sensorimotor integration during attentive speech perception.

Introduction

Speech is supported by sensorimotor integration, a (bidirectional) translation of its auditory and motoric signals. These translations, which occur during speech perception (Hickok and Poeppel, 2007; Evans et al., 2016; Schomers and Pulvermüller, 2016), can result in the cortical representation of articulatory features of speech, such as place of articulation, manner of articulation, and voicing. Particularly, the cortical representation of place of articulation features has been reported in dorsal speech regions, including motor and premotor areas (Pulvermuller et al., 2006), and somatosensory and supramarginal regions under passive listening (Correia et al., 2015). However, differential task requirements can modulate cortical representations of articulatory features. For example, variation of somatotopic activations in motor areas were found during passive sound perception involving different articulators (Pulvermuller et al., 2006a). Other researchers have reported differential patterns in superior temporal but not in the motor cortex during an incidental task with phonemes (Arsenault and Buchsbaum, 2015). Several reasons have been discussed to account for this variability; among them, differences in task demands across studies (e.g., type of task, number of items, but also passive vs. active tasks, i.e., selective attention) as well as third factors such a covariation of manner of articulation during place processing.

Although the underlying neural mechanisms of speech perception and attention remain elusive, they are explained in terms of network dynamics (Friederici and Singer, 2015) and, in particular, of theta-gamma amplitude or phase coupling (Giraud and Poeppel, 2012; Hyafil et al., 2015) of neural activity. For example, attention can implement phase resetting and entrainment of neuronal oscillations to a relevant stimulus stream (Lakatos et al., 2008) and can rapidly change the spectrotemporal receptive field to enhance task-relevant stimulus properties (Fritz et al., 2003). Importantly, selective attention has been shown to generate

93 spatial coupling patterns between prefrontal and feature-specific cortical areas (Baldauf and
94 Desimone, 2014).

95 Ultra high-field 7 Tesla fMRI allows investigating the living human brain with unprecedented
96 high spatial resolution (Yacoub et al., 2001), signal-to-noise ratio (Vaughan et al., 2001), and
97 specificity (Polimeni et al., 2010). Beyond measurement improvements, multivariate pattern
98 analysis (MVPA) further increases the sensitivity of experimental contrasts by exploiting
99 concurrent spatial patterns of fMRI responses (Norman et al., 2006). Crucially, MVPA has
100 allowed unraveling information representation of abstract speech features (Correia et al.,
101 2015; Evans and Davis, 2015).

102 Here, we aim to minimize the effect of covariation in manner of articulation during place of
103 articualtion processing by generalization across manner. We study the cortical representation
104 of place of articulation features of syllables using 7T fMRI and MVPA-based cross-decoding
105 (i.e., generalization). Specifically, we exploited the acoustic variation imposed by two
106 manners of articulation (e.g., stop and fricative) to identify patterns discriminative of place of
107 articulation (i.e., bilabial/labiodental versus alveolar) and tested these patterns in a third
108 (unseen) manner of articulation (e.g., nasal). This procedure capitalized on the acoustic
109 variation imposed by different manner of articulation features to extract specific patterns for
110 place of articulation features (Ladefoged and Johnson, 2010).

111 This generalization was applied to two different tasks with identical auditory stimuli (i.e.,
112 attention to consonants and attention to vowels) to investigate the neural representation of
113 place of articulation features during selective attention to speech and the possible effects of
114 attention on the neural representation. Previous studies have shown representation of place of
115 articulation features in sensorimotor regions (Correia et al., 2015) and a modulatory role of
116 attention in sound and phonetic representations of speech (Bonte et al., 2014; Downer et al.,

2015). Based on these studies, we used decoding of place of articulation features as a metric of sensorimotor processes during speech perception. We then investigated whether decoding was possible on either task and, if so, whether sensorimotor integration is modulated by the selective attention to different syllable components (i.e., vowel or consonants). The idea behind choosing vowels and consonants tasks was that place of articulation is more representative for consonant identification, and less for vowel identification in Consonant-Vowel structures. This orthogonal stimuli-task arrangement allowed assessing the effect of attention on the place of articulation representation without explicit task focus to place.

Materials and Methods

Participants

Fourteen native Dutch speakers (mean ± SD age, 27.2 ± 5.1 years; 9 females; 2 left-handed) underwent blood oxygenation level-dependent (BOLD) signal fMRI scanning while performing two delayed-match-to-sample tasks. Participants had no history of neurological or systemic diseases and reported normal hearing abilities. All participants received monetary compensation for their participation and signed written informed consent. The Ethical Committee of the authors' Home Institution approved the study.

Stimuli

Stimuli were 18 consonant-vowel (CV) syllables recorded by 3 female native Dutch speakers, generating 54 unique tokens. The syllables were the result of the CV combinations of six consonants (p, t, f, s, m, n) and three vowels (a, i, u). This subset of consonants was selected because it allowed us to build a balanced stimuli matrix that covers two features of place of articulation (i.e., bilabial/labiodental and alveolar) and three features of manner of articulation (i.e., stop, fricative, and nasal). This arrangement was needed to perform MVPA-based cross-decoding analysis and to keep the necessary number of trials for single-trial classification (see

Multivariate analysis below). Syllable recordings were selected from a subset of recordings previously used in our laboratory (Correia et al., 2015). Briefly, the consonants composing the syllables for every articulatory feature were bilabial/labiodental, ('p', 'f', 'm') and alveolar ('t', 's', 'n') for place of articulation; and stop ('p', 't'), fricative ('f', 's'), and nasal ('m', 'n') for manner of articulation (Figure 1A). The three different vowels and three different speakers introduced acoustic variability useful for classification. All stimuli were recorded in a soundproof chamber at a sampling rate of 44.1 kHz and digital-to-analog converted with 16-bit resolution. Stimuli were presented in the MRI scanner via MR-compatible earphones with a linear frequency transfer of up to 8 kHz (Model S14, Sensimetrics Corporation, USA). Before starting the experiment, the volume was adapted to each subject's audible and comfortable perceptual level.

Experimental design

Participants performed two delayed-match-to-sample tasks in four runs of 54 trials each (i.e., each syllable token was presented once per run) during fMRI acquisition. In cognitive terms, this task allowed paying attention to a certain aspect of the acoustic stimuli – from now on referred to as "attend to vowels" and "attend to consonants." The task also involved remembering and matching the attended vowel or consonant later on with a visually presented token (i.e., a written vowel or consonant) of the same category (with 50% match/mismatch response proportions). Each run was divided into two blocks of 27 trials, one for each task, and the tasks where counterbalanced across runs and participants. During the attend to vowels task, participants heard a syllable and received, after 6-8 seconds, a written vowel as "match cue." Participants were instructed to match the vowel of an auditorily presented CV syllable with a written vowel cue as fast and accurately as possible by pressing a button with the right index ('match') or middle ('mismatch') finger. During the attend to consonants task, participants heard the same syllables (of the attend to vowels task) and received a written

consonant as cue that should be matched with the auditory syllable (see Figure 1B). Importantly, the features of interest (i.e., place of articulation) were never part of an explicit task to avoid confounds in attentional demands. The delay-match-to-sample task is widely used to precisely control the subject's attention. Thus, data were analyzed only for the first part of each trial before the cue onset to avoid that relevant signal was contaminated with matching and response related processing period signals.

The task was clearly explained to each subject outside the scanner, and they received at the beginning of each block an introduction display indicating the syllable component they should attend to (i.e., vowels or consonants) at the beginning of each task. A trial consisted of (1) a speech stimulus sound presentation (i.e., a CV syllable token of 340 milliseconds duration) followed, after an inter-stimulus interval (ISI) of 6-8 seconds, by (2) a written cue (i.e., a consonant or vowel letter matching or not the sound token depending on the specific block and trial of 2 seconds duration, written in Times New Roman, font size 30, black color); and then followed by (3) the subject's 'match'/'mismatch' immediate response. The sound syllables and the visual cue (i.e., a written vowel or consonant) matched in 50% of the trials, and 'match'/'mismatch' trials were balanced across attention conditions and randomized per subject. Mismatch cues were always of the same category (i.e., vowels in the attend to vowels task and consonants in the attend to consonants task). In total, each run lasted 15 minutes and the behavioral responses were collected through an MR-compatible button box (Current Designs, 8-button response device, HHSC-2x4-C; Philadelphia, USA).

We used fast sparse image acquisition to have a 500 ms silent gap to present each syllable sound (Di Salle et al., 2003). The fMRI acquisition was set up as a slow event-related design. The inter-trial interval (ITI) between consecutive auditory stimuli was relatively long (i.e., 14 seconds on average; range: 12 to 16 seconds) to allow independent BOLD signal estimation per trial (see Figure 1C). The interval used to estimate the fMRI response to the spoken

191 syllable perception per trial was restricted to the first 6 seconds following the sound
192 presentation to avoid contamination from the processing of the visual cue, matching, or button
193 press. Moreover, as the acoustic, phonetic, and phonological features of the presented
194 syllables were identical across trials and attention conditions, the effects on the cortical
195 representations pertain to differences in attention. All stimuli, event identities, and timings
196 were presented and logged using Presentation from Neurobehavioral Systems
197 (www.neurobs.com, RRID: SCR_002521).

198 *Functional MRI acquisition*

199 Functional and anatomical volumes were acquired on a whole-body Siemens Magnetom 7
200 Tesla scanner (Siemens, Erlangen, Germany) and a 32-channel head-coil (Nova Medical Inc.;
201 Wilmington, USA) at the author Home Institution Imaging Center. For all functional runs, we
202 acquired whole brain high-resolution accelerated multiband gradient echo (T2*-weighted)
203 echo-planar imaging (EPI; Moeller et al., 2010; Setsompop et al., 2012) data (echo time, TE =
204 21 ms; repetition time, TR = 2,000 ms; time of acquisition, TA = 1,500; delay in TR (silent
205 gap) = 500 ms; multi-band factor = 3; generalized auto-calibrating partially parallel
206 acquisitions (GRAPPA) g-factor = 2; flip angle = 72º; field of view, FOV = 198 mm; voxel
207 size = 1.5 x 1.5 x 1.5 mm^3; number of slices = 72, without gap between slices) for each
208 participant. To correct for EPI geometric distortions, 10 volumes with opposite phase
209 encoding directions (i.e., posterior to anterior and anterior to posterior) were additionally
210 acquired using the same acquisition parameters as in the functional runs. After the first 2
211 functional runs, we acquired a tridimensional T1-weighted magnetization prepared rapid
212 acquisition gradient echo (3D-MP2RAGE; Marques et al., 2010) volume (240 sagittal slices;
213 voxel size = 0.65 x 0.65 x 0.65 mm^3; first inversion time TI1 = 900 ms; second inversion time
214 TI2 = 2,750 ms; TE = 2.51 ms; TR = 5,000 ms; first nominal flip angle = 5º; second nominal
215 flip angle = 3º) per participant.

Functional MRI preprocessing

Anatomical and functional data were analyzed using BrainVoyager QX (version 2.8.4; Brain Innovation; Maastricht, Netherlands, RRID: SCR_013057), and custom code written in MATLAB (R2014a version 8.3.0.532; The Mathworks Inc.; Natick, MA, USA, RRID: SCR_001622). Anatomical images were interpolated to a nominal voxel size of 1.5 x 1.5 x 1.5 mm^3 matching the functional images' resolution. The functional images were corrected for motion artifacts using the 3D rigid body motion correction algorithm implemented in BrainVoyager QX and all functional runs were aligned to the first volume of the second functional run. We corrected the EPI distortions using the *topup* tool implemented in FSL (RRID: SCR_002823, (Smith et al., 2004). The reversed phase encoding images, acquired after the anatomical images, were used to estimate the susceptibility-induced off-resonance field and, then, to correct the distortions in the remaining functional runs. After this correction, functional data were high-pass filtered using a general linear model (GLM) Fourier basis set of eleven cycles sine/cosine, including a linear trend removal. Functional volumes per run were co-registered and aligned to the anatomical scan using rigid body transformations (i.e., six parameters: three translations and three rotations). Finally, functional images were normalized by transformation into Talairach space (Jean and Tournoux, 1988).

Anatomical segmentation and cortex-based alignment. Intensity inhomogeneities in T1-weighted images were first corrected using Statistical Parametric Mapping 12 (SPM12, RRID: SCR_007037 Ashburner and Friston, 2005) software. The resulting images were used to perform volumetric segmentation with the FreeSurfer analysis software (version 5.3.0, http://surfer.nmr.mgh.harvard.edu/, RRID: SCR_001847) (Dale et al., 1999). Briefly, this processing includes motion correction, removal of non-brain tissue using a hybrid watershed/surface deformation procedure, automated Talairach transformation, intensity normalization, tessellation of the gray/white matter boundary, automated topology correction,

and surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class. Quality control was performed by visually inspecting each subject's brain after the process was finished. Remaining errors were manually corrected using ITK-SNAP software (version 3.4.0, www.itksnap.org, RRID: SCR_002010) (Yushkevich et al., 2006). The resulting binary maps were then used to reconstruct individual 3D meshes of the cortical surfaces and aligned using a moving target-group average approach based on curvature information (i.e., cortex-based alignment – CBA) to obtain an anatomically aligned group-averaged 3D surface representation of all the subjects (Goebel et al., 2006; Frost and Goebel, 2012). Functional data were analyzed (see Multivariate analysis below) in the volume space and then projected to the average surface, to perform group statistics and visualization in the aligned CBA space.

Univariate analysis

Single subject GLM analysis was performed on fMRI signal time courses normalized with percentage transform in volume space. Next, they were mapped onto surface space by sampling the values located between 1 mm below the gray/white matter boundary and up to 3 mm into the gray matter towards the pial surface using trilinear interpolation and averaging. This sampling resulted in a single value per vertex in the subject's cortex mesh, and then the values were aligned to the cortical group surface mesh using CBA. Random-effects GLM analysis was performed on these individual time course data. The conditions were collapsed across the speaker and vowel dimensions, thus yielding 12 predictors, 6 predictors for each consonant times 2 tasks. Each predictor was convolved with a canonical double gamma hemodynamic response function (HRF). Functional maps (i.e., average β values) were calculated to assess sound-evoked fMRI responses during attend to vowels and attend to consonants tasks (i.e., all sounds attend to vowels task > baseline; all sounds attend to

consonants task > baseline). Task differences were analyzed for attend to consonants and attend to vowels task activity (i.e., attend to consonants task > attend to vowels task). Univariate stimulus effects were analyzed for each place of articulation feature independently (e.g., p_Con + f_Con + m_Con + p_Vow + f_Vow + m_Vow > t_Con + s_Con + n_Con + t_Vow + s_Vow + n_Vow). All functional contrast maps were corrected for multiple comparisons by applying a permutation-based cluster-size threshold with an initial cluster forming threshold of $p = 0.05$. The cluster-size threshold was based on the distribution of maximum cluster sizes obtained in the 2,000 permutations, only considering clusters whose size was larger than the 95% quantile.

Multivariate analysis

To investigate the cortical representations of the different articulatory features, we used MVPA in combination with a moving volumetric searchlight approach (Kriegeskorte et al., 2006). The purpose of the multivariate analysis was to decode articulatory features for each syllable independent from their specific phonetic signatures or acoustical properties. Therefore, we used a classification approach based on cross-decoding, generalizing place of articulation features across different dimensions of manner of articulation. Specifically, we trained a classifier to discriminate bilabial/labiodental vs. alveolar places of articulation using syllables exhibiting two manners of articulation dimensions (e.g., stop and fricative) and tested whether this training was transferable to syllables exhibiting a third (unseen) manner dimension (e.g., nasal). MVPA generalization analysis was performed within each subject using all three combinations of generalization (i.e., leaving one manner of articulation out for testing at each split). The obtained averaged accuracies were submitted to a group analysis using random permutations for significance and nonparametric permutation-based cluster-size thresholding for multiple comparisons correction (see Statistical analysis – Group statistics below).

Within-subject decoding

Before classification, BOLD time courses were selected by extracting the responses in each trial and each cortical voxel. The BOLD responses were estimated by fitting a standard HRF using multiple linear regression, taking the first three samples per trial after sound onset (i.e., corresponding to 6,000 ms). The regression coefficients resulting from the fitting of single trial data in the cortical mask were used to build an fMRI feature space (i.e., number of trials by number of cortical voxels), which was then used in the multivariate decoding. During the multivariate decoding, we kept trials belonging to each task separate.

We limited our analysis to gray matter voxels using a subject-specific cortical ribbon mask based on intensity values in the T1-weighted images (i.e., cortical ribbon segmentation). We constrained our analysis to the cortical ribbon, as we were mainly interested in cortical processing. Ribbons allowed us to exclude volumes containing white matter and subcortical voxels from the analysis, which increased the number of dimensions. For the searchlight analysis, a sphere with radius of 5.5 voxels (i.e., 8.25 mm) was moved through the cortical ribbon, which defined a feature space of 65,000 voxels on average. In each searchlight, we performed generalization of place of articulation across manner of articulation using Linear Discriminant Analysis (LDA) with diagonal covariance matrix. Given the large number of searchlights, we employed a fast Matlab implementation described in Ontivero-Ortega et al. (2017).

We also investigated the cortical representation of place of articulation independent of the task effect. We did so by performing classification and generalization across manner of articulation in the pooled data from the two tasks (i.e., attend to vowels and attend to consonants). By doing this we use twice the amount of data as in the previous analyses, and can expect more robust and reliable findings due to the power increase in single subject

315 analyses. To avoid any task bias, we controlled for the number of trials belonging to each task
316 in each cross-validation split maintaining a balanced contribution of each task during the
317 training and testing procedures. The decoding was performed as described above.

318 *Multivariate statistical analysis – Group statistics*

319 In each subject, classification accuracies were obtained, separately for each task (i.e., attention
320 to consonant and attention to vowel), for each searchlight centered on voxels belonging to the
321 subject's specific cortical ribbon. To perform a group analysis, these accuracies were
322 projected onto each subject's reconstructed cortical surface and then mapped back onto the
323 group average cortical sheet using CBA. After the re-alignment, we considered in the group
324 analysis only the vertices common to all subjects. To assess uncorrected significance of the
325 classifications at the group level, we considered a non-parametric permutation test at each
326 searchlight location independently. For each vertex, we took all the subject accuracies and the
327 associated mean, and, using a resampling approach, we estimated the probability that such
328 mean or a higher value could occur if the data were obtained under the null hypothesis (H0)
329 that the decoding population mean is at chance level (i.e., 50%). This probability is the p-
330 value associated with the observed group mean. The resampling strategy was based on the
331 fact that, under H0, the likelihood of the observed accuracies is symmetric around chance
332 (i.e., 55% and 65% are equally likely if the population mean is 50%), and we can, therefore,
333 build many datasets by switching some of the subjects' accuracies around chance (Good,
334 2005, section 3.2.1). We ran 2,000 random permutations (Monte Carlo permutations) and for
335 each searchlight we determined the p-value as the ratio between how many times the mean of
336 a population resample equaled or exceeded the observed mean and the total number of
337 permutations (adding 1 to both numerator and denominator to avoid 0 p-values, hence the
338 lowest p-value was 1/2001).

Finally, to correct for multiple comparisons, we employed a permutation-based cluster-size thresholding. We considered an initial Cluster Forming Threshold equal to 0.05 and, then, for each permutation, we tagged as significant those vertices whose observed mean was higher than the 1- quantile (obtained with the resampling procedure described above). For each of these "false positive" maps, we determined the maximum extent of clusters (on the surface) and built a distribution of cluster sizes across the 2,000 permutations. Clusters of significant vertices, or equal to 0.05, in the observed data that were larger than the 95% quantile of the obtained distribution were deemed significant and corrected for multiple comparisons. It is worth mentioning that this procedure is based on a permutation strategy that does not suffer from inflation of false positive rate, as recently shown in Eklund et al. (2016).

Results

Behavioral results in fMRI experiment

To test whether there were any significant difference in reaction times, a factorial analysis of variance (ANOVA) was conducted, with task (attention to vowels and attention to consonants), place of articulation (bilabial/labiodental and alveolar), manner of articulation (stop, fricative, and nasal), speaker (speaker 1, speaker 2, and speaker 3), vowel (a, i, and u), and subject (subject 1 to 14, treated as random factor) as main factors on correct trials only. There was no effect across conditions in behavior. Reaction times for the vowel (mean = 843 ms, SD = 402) and the consonant task (mean = 820 ms, SD = 331) did not significantly differ [$F(1, 2541) = 1.26$, $p = 0.261$]. There were no significant main effects or interactions (all p-values > 0.163).

Functional MRI univariate results during the attend to consonants and attend to vowels tasks

During both the attend to vowels and attend to consonants tasks, the stimuli evoked significant BOLD responses in an extensive area of the superior temporal cortex,

363 encompassing early auditory areas (i.e., Heschl's gyrus and Heshl's sulcus), the planum
364 temporale, the superior temporal sulcus (STS) and gyrus (STG), and the posterior part of the
365 middle temporal gyrus (MTG). Additionally, the insula, supramarginal gyrus, intraparietal
366 sulcus and lateral prefrontal cortex including inferior frontal gyrus (IFG) were activated
367 during both tasks (Figure 2). Finally, univariate analyses did not reveal main effects of task or
368 main effect of place of articulation.

369 *Functional MRI decoding results*

370 To investigate the cortical representation of place of articulation features during the two
371 attentional conditions (i.e., attend to consonants and attend to vowels), we implemented a
372 classification method that relied on generalizing the discriminability of two places of
373 articulation features (i.e., bilabial/labiodental and alveolar) across variation of three manner of
374 articulation features (i.e., stop, fricative, and nasal).

375 Generalization maps after correction for multiple comparisons using cluster-size thresholding
376 (p-clust < 0.05) during the performance of the attend to vowels task are presented in Figure
377 3A. The generalization maps revealed successful decoding of place of articulation within
378 regions of the brain's language network, bilaterally. In the left hemisphere, clusters were
379 distributed across different regions including posterior temporal, temporo-parietal, insular,
380 anterior infero-frontal, frontal, and premotor medial regions. Specific regions included the
381 posterior superior temporal sulcus (STS), supramarginal gyrus (SMG), temporoparietal
382 junction, supplementary motor area (SMA), anterior insula, and anterior portion of the inferior
383 frontal gyrus (aIFG – pars orbitalis and pars triangularis). In the right hemisphere, clusters
384 included superior temporal, insular, inferior-motor and inferior-frontal regions. Specific
385 regions encompassed mid-posterior superior temporal plane (mSTG and pSTG), including
386 Heschl's Gyrus (HG), inferior central sulcus, subcentral gyrus, anterior insula, and posterior

(pars opercularis) and anterior (pars triangularis and orbitalis) parts of the inferior frontal gyrus (IFG).

Generalization maps for place of articulation during the performance of the attend to consonants task revealed significant clusters in motor regions, insula, and anterior frontal regions in the right hemisphere only (see Figure 3B). More specifically, these clusters included the anterior-superior portion of the angular gyrus (AG), inferior precentral gyrus, inferior central sulcus, subcentral gyrus, anterior-superior insula, posterior IFG, superior frontal sulcus, and anterior middle frontal gyrus.

To test for the modulation of place of articulation feature representations between tasks, we compared the classification accuracies of these representations within each searchlight across the two tasks using a two-sample permutation test. This analysis yielded no significant clusters with correction for cluster-size in any of the hemispheres.

To explore possible tendencies in the data beyond the rigorous cluster-size thresholding correction, we examined the maps without correction (i.e., $p < 0.05$ uncorrected). We observed that the left hemisphere had more information (clusters) than the right hemisphere. Particularly, we found a large continuous cluster of 134 square millimeters located in the left anterior insula. For this cluster, the classification of place of articulation features during the attend to vowels task exhibited higher accuracies than the classification of place of articulation features during the attend to consonants task (see white outline in Figure 3 and Figure 4). Aditionally, the posterior IFG bilaterally tended to show larger classification during attend to vowels than during attend to consonants task (see Figure 4).

To increase the power in decoding at the single subject level, we renounced to task specificity by decoding the representation of place of articulation independently of the task effects. We

pooled together all the trials from the attend to vowels and attend to consonants tasks and performed the classification of place of articulation ignoring whether a trial belonged to the attend to vowels or attend to consonants task. Using twice as many data as in the task-specific analysis, we expected this classification to be more robust and hence yield a higher number of significant clusters. We found bilateral clusters in the angular, supramarginal, and inferior frontal (i.e., pars opercularis and triangularis) gyri, the anterior insula, middle frontal rostral areas, middle portion of the superior frontal and middle frontocaudal areas, and the intraparietal sulcus. We also found clusters in the right posterior portion of the middle temporal gyrus, the left precentral gyrus, and the right subcentral gyrus (see Figure 5).

Discussion

The present study investigated the spatial representation of place of articulation features during the attentive perception of spoken consonant-vowel syllables (i.e., attend to vowels and attend to consonants). Using identical acoustic input and generalization across variation due to manner of articulation, we provided evidence for cortical representation of place of articulation features. The generalization across manner of articulation allowed us to maximize the acoustic invariance in fMRI classification (Correia et al., 2015) to counteract the differences in acoustic properties (e.g., the variance of the second formant frequency of the vowel segment due to perseverative coarticulation) that often accompany those in place of articulation (Ladefoged and Johnson, 2010).We found representation of place of articulation in separate tasks as well as in the pooled set (i.e., with both attention tasks). During the attend to vowels task, generalization maps of place of articulation features indicated feature sensitivity within bilateral clusters in superior and posterior temporal, insular, and prefrontal regions. During the attend to consonants task, place of articulation features were represented in temporoparietal, insular, and frontal regions within the right hemisphere only. The representation of place of articulation features independent of task effects (i.e., performing the

435 generalization analysis in the pooled data) showed similar clusters to those obtained with each
436 task separately. In addition, this analysis yielded bilateral effects as well as a more prominent
437 contribution of frontal regions.

438 We observed that the brain represents the different features of place of articulation (i.e.,
439 bilabial/labiodental and alveolar) of speech sounds during attentive listening, as shown by
440 generalization across manner of articulation. This observation supports the relevance for place
441 in the context of sensorimotor representation of speech perception (Pulvermuller et al., 2006a;
442 Correia et al., 2015). The present result further supports previous findings that report within-
443 place feature differentiations in STS without generalization across manner (Davis and
444 Johnsrude, 2003; Scott and Johnsrude, 2003; Eulitz and Lahiri, 2004).

445 From a descriptive point of view, the cortical representation included temporal and frontal
446 regions during the attend to vowels task, and right middle and inferior frontal regions during
447 the attend to consonants task. This observation can be discussed in the light of differences in
448 (a) phonological feature specification for vowels and consonants, (b) stimulus processing for
449 consonants and vowels over time, and (c) attentional processing across the stimuli.
450 Consonants belonged to a *specific* feature class (i.e., bilabial/labidental or alveolar), whereas
451 vowels were equally represented across *all* syllables and place of articulation classes, see
452 Figure 1A (Eulitz and Lahiri, 2004; Scharinger et al., 2016). The temporal processing of
453 consonants and vowels in a CV structure could also have influenced the spatial distribution of
454 the cortical representation; it has been long recognized that the syllable onset has the most
455 relevant status in speech comprehension, whereas stimulus endings can be ignored more
456 easily (Marslen-Wilson and Zwitserlood, 1989).

457 The observed cortical representation could also have related to attentional processing and
458 variation of relevance-based selection across the two tasks. Attending to vowels meant

459 expecting the target to occur always during the *last* part of the syllable. However, although the
460 information related to the consonants is irrelevant for the task, preceding (irrelevant) linguistic
461 information cannot be ignored, as has been shown by studies on subliminal priming (Schütz et
462 al., 2007) and phonological priming (Bles and Jansma, 2008). Therefore, consonant-related
463 information could have still been linguistically processed during attention to vowels, thereby
464 impacting the respective representations within the language system. Attending to consonants
465 meant, in turn, expecting the target to occur always during the *first* part of the syllable, so the
466 information presented after the target (i.e., the vowel) could be easily ignored or discarded.
467 Consequently, the cortical representation of place of articulation features during the attend to
468 vowels task was found to include all areas initially expected for the attend to consonants task
469 (e.g., Correia et al., 2015). Similarly, consonant to vowel transition features may have been
470 amplified during the attend to vowels task, involving specific bilateral temporal regions
471 (Humphries et al., 2014). As during the attend to consonants task no transition (to a
472 subsequent vowel) should be amplified, brain regions dealing with such transitions were not
473 engaged. Future studies could also include other sequences of syllable components (e.g.,
474 Vowel-Consonant syllables or Consonant-Vowel-Consonant structures) to clarify the effect of
475 the relevance-based selection.

476 An alternative possibility pertains to general-feature sensitivity of the language system during
477 attention to vowels and specific-feature selectivity during attention to consonants, as both
478 syllable components can exhibit acoustic and articulatory properties (Ladefoged and Johnson,
479 2010). Given their relatively clearer acoustic features (Narayan et al., 2010), vowels could
480 have had relatively higher perceptual (i.e., auditory) saliency in comparison with consonants,
481 thus explaining the extended representation of place of articulation features during the
482 attention to vowels task. However, this possibility is unlikely because our participants did not

differ in reaction times between tasks, and higher saliency should result in faster reaction times during detection and match to sample tasks.

The comparison between the two classification maps of place of articulation features yielded a large continuous cluster (uncorrected) in the left anterior insula for attention to vowels vs. attention to consonants. This result points towards a role of the anterior insula in the modulation of the representation of place of articulation, determined by selective attention to vowels in CV syllables. A relevance-based selection, resulting from the specificity of our design, might have dictated a task set where attention to vowels required relatively greater demands on top-down selectivity (i.e., in the attempt to ignore the first part of the syllable). This interpretation is in agreement with the role of the anterior insula as part of a task-set system (Dosenbach et al., 2006) in volitional top-down control (Nelson et al., 2010) and alertness (Sadaghiani et al., 2010; Clemens et al., 2011). Moreover, the anterior insula has been incorporated within the hierarchy of the language network (Davis and Johnsrude, 2003), for example with a role in articulatory planning (Dronkers, 1996; Baldo et al., 2011).

The lack of significant clusters representing place of articulation features in other parts of the temporal lobe (e.g., primary auditory cortex) was expected given our analysis approach. The main purpose of combining MVPA with generalization was to maximize the extraction of the information about the abstract (higher order) features under study (i.e., place of articulation) while minimizing the impact of variation in *acoustic* (i.e., manner of articulation) information. Thus, the clusters show where information about place of articulation is represented, free from perceptual properties and information processing of the sounds. Moreover, the univariate analysis results support a critical role of the auditory cortices in speech perception and auditory processing (see Figure 2). However, it should be noted that the univariate maps are showing the main effect to sound compared to silence (i.e., overall relative changes of activation in cortical regions in response to sound). Therefore, these maps reflect sound

508 responsive areas (to speech in our experiment), rather than isolated language-specific areas.
509 The representation of place of articulation revealed by the generalization analysis in separate
510 tasks was also confirmed when the trials from the two tasks were pooled. The additional
511 clusters in frontal areas could have resulted from both the increase in the statistical power
512 (i.e., using twice the number of trials in comparison to the task-specific analyses) and the
513 higher sensitivity of multivariate pattern analysis (i.e, compared to univariate analysis;
514 (Kriegeskorte et al., 2006; Norman et al., 2006; Mur et al., 2008).

515 Overall, our generalization approach showed a network of distributed brain regions (including
516 frontal and sensorimotor areas), where articulatory features are represented during the
517 perception, attention, and short-term memory storage within the delay-match-to-sample tasks.
518 This finding fits well with the cortical areas critical for language comprehension (e.g., the
519 angular and supramarginal gyri; Turken and Dronkers, 2011), sensorimotor integration
520 (Pulvermuller et al., 2006b; Hickok et al., 2011; Schomers and Pulvermüller, 2016), and
521 semantic knowledge retrieval (e.g., angular gyrus; Dronkers et al., 2004; Binder et al., 2009).
522 Moreover, lesion studies showing sentence comprehension impairment have also suggested a
523 role for dorsolateral frontal areas in auditory information integration (e.g., short-term verbal
524 memory rehearsal during sentence comprehension: Smith and Jonides, 1999; and the
525 representation of speech sequences: Fuster, 2001).

526 Considering the observed representation of place of articulation, it would seem reasonable to
527 analyze the individual phonological features in an independent manner. However, we could
528 not do so within our dataset due to the setup of the current experiment, in which the total
529 number of trials was limited. Limitations were related to the sparse sampling acquisition
530 necessary to avoid scanner noise interference, which increased the trial duration. In addition,
531 stimuli tokens were reduced to a subset required to introduce enough variance for
532 classification. These choices, though necessary for the purpose of the study, resulted in

533 selection of material with few individual place of articulation features. Previous studies have
534 also shown representation of individual abstract phonological features (Obleser et al., 2003,
535 2004), notions of underspecification (Scharinger et al., 2016), as well as a multidimensional
536 space of features (i.e., spectral peak or voice onset time) in the encoding of acoustic
537 parameters of speech (Mesgarani et al., 2008, 2014).

538 In summary, here we showed classification of place of articulation features generalized across
539 manner of articulation for the first time. Our results provide evidence for cortical
540 representation of place of articulation features during attentive syllable perception (i.e.,
541 attention to the different syllable components) in sensorimotor, temporoparietal, and frontal
542 areas within the language network. This cortical representation was revealed in additional
543 clusters (e.g., bilateral frontal and sensorimotor areas) when increasing the statistical power.
544 Additionally, we observed a more bilateral distribution for attend to vowels, a more unilateral
545 distribution for attend to consonants, and a trend towards a modulatory role of the left anterior
546 insula in selective attention of speech sounds. To conclude, these data support sensorimotor
547 integration during attentive speech perception and demonstrate that a generalization approach
548 can be used to exclude 'common factors', such as perceptual properties, from the analysis.

549 **References**

550 Arsenault JS, Buchsbaum BR (2015) Distributed Neural Representations of Phonological
551 Features during Speech Perception. J Neurosci 35:634–642 Available at:
552 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2454-14.2015.

553 Ashburner J, Friston KJ (2005) Unified segmentation. Neuroimage 26:839--851 Available at:
554 http://dx.doi.org/10.1016%252?j.neuroimage.2005.02.018.

555 Baldauf D, Desimone R (2014) Neural Mechanisms of Object-Based Attention. Science (80-)
556 344:424–427 Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1247003.

557 Baldo J V, Wilkins DP, Ogar J, Willock S, Dronkers NF (2011) Role of the precentral gyrus
558 of the insula in complex articulation. Cortex 47:800–807 Available at:
559 http://dx.doi.org/10.1016/j.cortex.2010.07.001.

560 Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A
561 critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex
562 19:2767–2796.

563 Bles M, Jansma BM (2008) Phonological processing of ignored distractor pictures, an fMRI
564 investigation. BMC Neurosci 9:20 Available at:
565 http://bmcneurosci.biomedcentral.com/articles/10.1186/1471-2202-9-20.

566 Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014) Task-Dependent Decoding
567 of Speaker and Vowel Identity from Auditory Cortical Response Patterns. J Neurosci
568 34:4548–4557 Available at:
569 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4339-13.2014.

570 Clemens B, Zvyagintsev M, Sack AT, Sack A, Heinecke A, Willmes K, Sturm W (2011)
571 Revealing the functional neuroanatomy of intrinsic alertness using fMRI: methodological
572 peculiarities. PLoS One 6:e25453 Available at:
573 http://www.ncbi.nlm.nih.gov/pubmed/21984928.

574 Correia JM, Jansma BM, Bonte M (2015) Decoding Articulatory Features from fMRI
575 Responses in Dorsal Speech Regions. J Neurosci 35:15015–15025 Available at:
576 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0977-15.2015.

577 Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and
578 surface reconstruction. Neuroimage 9:179–194 Available at:
579 http://www.ncbi.nlm.nih.gov/pubmed/9931268.

580 Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J
581 Neurosci 23:3423–3431.

582 Di Salle F, Esposito F, Scarabino T, Formisano E, Marciano E, Saulino C, Cirillo S, Elefante
583 R, Scheffler K, Seifritz E (2003) fMRI of the auditory system: understanding the neural

584 basis of auditory gestalt. Magn Reson Imaging 21:1213–1224 Available at:
585 http://www.ncbi.nlm.nih.gov/pubmed/14725929.

586 Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund
587 ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the
588 implementation of task sets. Neuron 50:799–812 Available at:
589 http://www.ncbi.nlm.nih.gov/pubmed/16731517.

590 Downer JD, Niwa M, Sutter ML (2015) Task Engagement Selectively Modulates Neural
591 Correlations in Primary Auditory Cortex. J Neurosci 35:7565–7574 Available at:
592 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4094-14.2015.

593 Dronkers NF (1996) A new brain region for coordinating speech articulation. Nature
594 384:159–161 Available at: http://www.ncbi.nlm.nih.gov/pubmed/8906789.

595 Dronkers NF, Wilkins DP, Van Valin RD, Redfern BB, Jaeger JJ (2004) Lesion analysis of
596 the brain areas involved in language comprehension. Cognition 92:145–177.

597 Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial
598 extent have inflated false-positive rates. Proc Natl Acad Sci 113:7900–7905 Available at:
599 http://arxiv.org/abs/1606.08199.

600 Eulitz C, Lahiri A (2004) Neurobiological evidence for abstract phonological representations
601 in the mental lexicon during speech recognition. J Cogn Neurosci 16:577–583.

602 Evans S, Davis MH (2015) Hierarchical organization of auditory and motor representations in
603 speech perception: Evidence from searchlight similarity analysis. Cereb Cortex 25:4772–
604 4788.

605 Evans S, McGettigan C, Agnew ZK, Rosen S, Scott SK (2016) Getting the Cocktail Party
606 Started: Masking Effects in Speech Perception. J Cogn Neurosci 28:483–500 Available
607 at: http://www.ncbi.nlm.nih.gov/pubmed/23647519.

608 Friederici AD, Singer W (2015) Grounding language processing on basic neurophysiological
609 principles. Trends Cogn Sci 19:329–338 Available at:
610 http://dx.doi.org/10.1016/j.tics.2015.03.012.

611 Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal
612 receptive fields in primary auditory cortex. Nat Neurosci 6:1216–1223 Available at:
613 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1671-15.2015.

614 Frost M, Goebel R (2012) Measuring structural-functional correspondence: Spatial variability
615 of specialised brain regions after macro-anatomical alignment. Neuroimage 59:1369–
616 1381.

617 Fuster JM (2001) The prefrontal cortex--an update: time is of the essence. Neuron 30:319–
618 333 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11394996.

619 Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging
620 computational principles and operations. Nat Neurosci 15:511–517 Available at:
621 http://dx.doi.org/10.1038/nn.3063.

622 Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest
623 (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group
624 general linear model analysis and self-organizing group independent component
625 analysis. Hum Brain Mapp 27:392–401 Available at:
626 http://doi.wiley.com/10.1002/hbm.20249.

627 Good PI (2005) Permutation, Parametric and Bootstrap Tests of Hypotheses. New York:
628 Springer-Verlag. Available at: http://link.springer.com/10.1007/b138696.

629 Hickok G, Houde J, Rong F (2011) Sensorimotor Integration in Speech Processing:
630 Computational Basis and Neural Organization. Neuron 69:407–422 Available at:
631 http://dx.doi.org/10.1016/j.neuron.2011.01.019.

632 Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev
633 Neurosci 8:393–402 Available at: http://www.nature.com/doifinder/10.1038/nrn2113.

634 Humphries C, Sabri M, Lewis K, Liebenthal E (2014) Hierarchical organization of speech
635 perception in human auditory cortex. Front Neurosci 8:1–12.

636 Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud AL (2015) Speech encoding by coupled
637 cortical theta and gamma oscillations. Elife 4:1–45.

638 Jean T, Tournoux P (1988) Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional
639 System: An Approach to Cerebral Imaging (Thieme Classics): J. Talairach:
640 9780865772939: Amazon.com: Books. Thieme:192 Available at:
641 http://www.amazon.com/Co-Planar-Stereotaxic-Atlas-Human-Brain/dp/0865772932.

642 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping.
643 Proc Natl Acad Sci U S A 103:3863–3868 Available at:
644 http://www.pnas.org/content/103/10/3863.abstract.

645 Ladefoged P, Johnson K (2010) A course in phonetics, 6th ed. Boston: Cengage Learning.

646 Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal
647 oscillations as a mechanism of attentional selection. Science 320:110–113.

648 Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele PF, Gruetter R (2010)
649 MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-
650 mapping at high field. Neuroimage 49:1271–1281 Available at:
651 http://dx.doi.org/10.1016/j.neuroimage.2009.10.002.

652 Marslen-Wilson W, Zwitserlood P (1989) Accessing spoken words: The importance of word
653 onsets. J Exp Psychol Hum Percept Perform 15:576–585 Available at:
654 http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.15.3.576.

655 Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in Human
656 Superior Temporal Gyrus. Science (80-) 343:1006–1010 Available at:
657 http://www.sciencemag.org/cgi/doi/10.1126/science.1245994.

658 Mesgarani N, David S V., Fritz JB, Shamma SA (2008) Phoneme representation and
659 classification in primary auditory cortex. J Acoust Soc Am 123:899 Available at:
660 http://scitation.aip.org/content/asa/journal/jasa/123/2/10.1121/1.2816572.

661 Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Uğurbil K (2010)
662 Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel
663 imaging with application to high spatial and temporal whole-brain FMRI. Magn Reson
664 Med 63:1144–1153.

665 Mur M, Bandettini PA, Kriegeskorte N (2008) Revealing representational content with
666 pattern-information fMRI--an introductory guide. Soc Cogn Affect Neurosci 4:101–109
667 Available at: http://scan.oxfordjournals.org/cgi/doi/10.1093/scan/nsn044.

668 Narayan CR, Werker JF, Beddor PS (2010) The interaction between acoustic salience and
669 language experience in developmental speech perception: Evidence from nasal place
670 discrimination. Dev Sci 13:407–420.

671 Nelson SM, Dosenbach NUF, Cohen AL, Wheeler ME, Schlaggar BL, Petersen SE (2010)
672 Role of the anterior insula in task-level control and focal attention. Brain Struct Funct
673 214:669–680 Available at:
674 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2886908&tool=pmcentrez&r
675 endertype=abstract.

676 Norman KA, Polyn SM, Detre GJ, Haxby J V (2006) Beyond mind-reading: multi-voxel
677 pattern analysis of fMRI data. Trends Cogn Sci 10:424–430 Available at:
678 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citat
679 ion&list_uids=16899397.

680 Obleser J, Eulitz C, Reetz H, Lahiri A (2003) Phonological features of speech segments are
681 reflected in the Auditory Evoked Brain Response around 100 ms post stimulus onset.
682 Brain:1643–1646.

683 Obleser J, Lahiri A, Eulitz C (2004) Magnetic Brain Response Mirrors Extraction of
684 Phonological Features from Spoken Vowels. J Cogn Neurosci 16:31–39 Available at:
685 http://www.mitpressjournals.org/doi/10.1162/089892904322755539.

686 Ontivero-Ortega M, Lage-Castellanos A, Valente G, Goebel R, Valdes-Sosa M (2017) Fast
687 Gaussian Naïve Bayes for searchlight classification analysis. Neuroimage:1–9 Available
688 at: http://linkinghub.elsevier.com/retrieve/pii/S1053811917307371.

689 Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an
690 imposed spatial activation pattern in human V1. Neuroimage 52:1334–1346 Available at:
691 http://dx.doi.org/10.1016/j.neuroimage.2010.05.005.

692 Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006a)
693 Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci 103:7865–
694 7870 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0509989103.

695 Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006b)
696 Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci 103:7865–
697 7870 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0509989103.

698 Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud A-L, Kleinschmidt A (2010)
699 Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous
700 electroencephalography/functional magnetic resonance imaging study. J Neurosci
701 30:10243–10250 Available at:
702 http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1004-10.2010.

703 Scharinger M, Domahs U, Klein E, Domahs F (2016) Mental representations of vowel
704 features asymmetrically modulate activity in superior temporal sulcus. Brain Lang
705 163:42–49 Available at: http://dx.doi.org/10.1016/j.bandl.2016.09.002.

706 Schomers MR, Pulvermüller F (2016) Is the Sensorimotor Cortex Relevant for Speech
707 Perception and Understanding? An Integrative Review. Front Hum Neurosci 10
708 Available at: http://journal.frontiersin.org/Article/10.3389/fnhum.2016.00435/abstract.

709 Schütz K, Schendzielarz I, Zwitserlood P, Vorberg D (2007) Nice wor_ if you can get the
710 wor_: Subliminal semantic and form priming in fragment completion. Conscious Cogn
711 16:520–532.

712 Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech
713 perception. Trends Neurosci 26:100–107 Available at:
714 http://www.sciencedirect.com/science/article/pii/S0166223602000371.

715 Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL (2012) Blipped-
716 controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging
717 with reduced g-factor penalty. Magn Reson Med 67:1210–1224.

718 Smith EE, Jonides J (1999) Storage and executive processes in the frontal lobes. Science
719 283:1657–1661 Available at:
720 http://www.sciencedirect.com/science/article/pii/095943889390204C.

721 Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-berg H,
722 Bannister PR, Luca M De, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J,
723 Zhang Y, Stefano N De, Brady JM, Matthews PM (2004) Advances in Functional and
724 Structural MR Image Analysis and Implementation as FSL Technical Report TR04SS2.
725 Neuroimage 23(S1):208–219.

726 Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension
727 network: converging evidence from lesion and connectivity analyses. Front Syst
728 Neurosci 5:1 Available at:
729 http://journal.frontiersin.org/article/10.3389/fnsys.2011.00001/abstract.

730 Vaughan JT, Garwood M, Collins CM, Liu W, Delabarre L, Adriany G, Andersen P, Merkle
731 H, Goebel R, Smith MB, Ugurbil K (2001) 7T vs. 4T: RF power, homogeneity, and
732 signal-to-noise comparison in head images. Magn Reson Med 46:24–30.

733 Yacoub E, Shmuel a, Pfeuffer J, Van De Moortele PF, Adriany G, Andersen P, Vaughan JT,
734 Merkle H, Ugurbil K, Hu X (2001) Imaging brain function in humans at 7 Tesla. Magn
735 Reson Med 45:588–594 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11283986.

736 Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided
737 3D active contour segmentation of anatomical structures: Significantly improved
738 efficiency and reliability. Neuroimage 31:1116–1128.

739

Legends

Figure 1.

Stimuli and experimental paradigm. A) Spoken stimuli matrix and articulatory properties. The 18 syllables were selected according to the place of articulation (i.e. bilabial/labiodental and alveolar) and manner of articulation (i.e. stop, fricative, and nasal). B) Experimental procedure and task description. Example of a typical match-to-sample trial during attend to vowels task. Subjects received instructions per block, in which they attended to consonants or vowels, respectively, and carried out a match-to-sample decision within slow-event related trials. Each block started with the visual presentation of a task cue (i.e., attention target vowel or consonant) indicating which task to perform in the next 27 trials. Each trial started with a fixation period in which a fixation cross was presented at the center of screen, together with a syllable sound (340 ms). After a jittered inter-stimulus interval (ISI jitter: 6 – 8 seconds) a visual cue (i.e., a written letter, vowel or consonant) was presented for 2 seconds, followed by the immediate subject's response by pressing a button either with the right index finger (for match trials) or middle index finger (for mismatch trials). The response was followed by a jittered ISI (4 – 10 seconds) to complete the jittered intertrial interval period (ITI: 12 – 16 seconds) before the next trial started. C) Schematic representation of the fMRI acquisition sequence and its relationship to the syllable sounds presented to the participants. Vow: attentd to vowels task, Con: attend to consonants task, Syll: syllable, TA: time of acquisition, TR: time for repetition, Phase: opposite phase encoding volumes acquired for distortion correction.

Figure 2.

Brain areas sensitive to speech sound processing during attend to vowels and attend to consonants tasks. Functional maps depicting the overall pattern of sound-evoked cortical responses during performance of A) attend to vowels (i.e., all sounds attend to vowels > baseline) and B) attend to consonants task (i.e., all sounds attend to consonants > baseline). P-value at cluster threshold (p-clust) < 0.05. Maps are visualized on inflated and aligned group-averaged representations of the left (LH) and right (RH) hemispheres of the fourteen subjects (light gray: gyri and dark gray: sulci). Color scale indicates average beta values.

Figure 3.

Cortical representation of place of articulation features, generalized across manner of articulation. Generalization maps depicting classification accuracies during A) attend to vowels and B) attend to consonants tasks. Insula cluster represents the outline (i.e., while line) of the largest continuous uncorrected cluster from the differences

768 between the two tasks. *P*-value at cluster-size threshold (p_{clust}) < 0.05. Searchlight radius 8.25 mm. Maps are
769 visualized on inflated and aligned group-averaged representations of the left (LH) and right (RH) hemispheres of
770 the fourteen subjects (light gray: gyri and dark gray: sulci). Color scale indicates classification accuracy (Acc)
771 percentages.

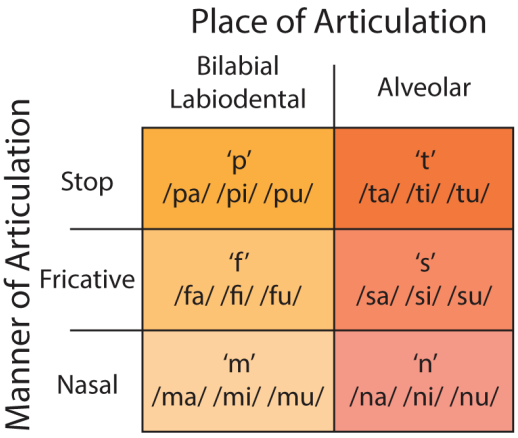
772 **Figure 4.**

773 Differences in classification of place of articulation features, generalized across manner of articulation. Maps
774 depicting the difference in the classification accuracies between attend to vowels and attend to consonants tasks.
775 The largest continuous cluster (i.e., area 134 square millimiters) was found in the left anterior insula, outlined in
776 white. *P*-value uncorrected (p_{uncorr}) < 0.05. Searchlight radius 8.25 mm. Maps are visualized on inflated and
777 aligned group-averaged representations of the left (LH) and right (RH) hemispheres of the fourteen subjects
778 (light gray: gyri and dark gray: sulci). Yellow–orange clusters show larger classification accuracies during the
779 attend to vowels task and green–blue clusters show lager classification accuracies during attend to consonants
780 task. Please note that the color scale does not directly relates to the colors used in Figure 3. Color scale indicates
781 differences in the classification accuracy (Acc) percentages.

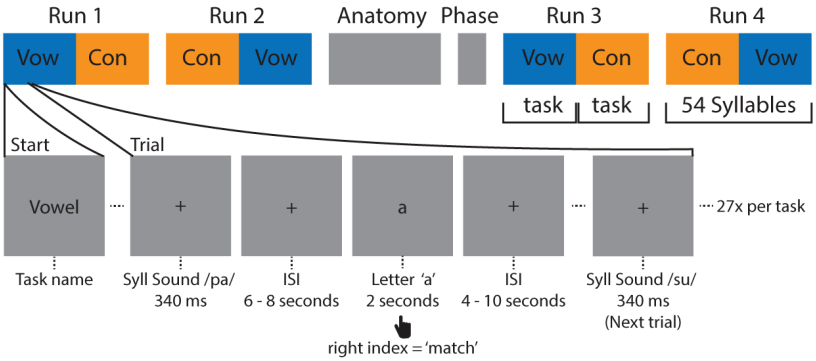
782 **Figure 5**

783 Cortical representation of place of articulation features, generalized across manner of articulation, pooling all
784 trails from two tasks. Generalization maps depicting classification accuracies calculated by pooling all trails
785 from attend to vowels and attend to consonants tasks. Corrected for multiple comparison with cluster-size
786 thresholding. Insula cluster represents the outline (i.e., white line) of the largest continuous uncorrected cluster
787 from the differences between the two tasks. *P*-value at cluster-size threshold (p_{clust}) < 0.05. Searchlight radius
788 8.25 mm. Maps are visualized on inflated and aligned group-averaged representations of the left (LH) and right
789 (RH) hemispheres of the fourteen subjects (light gray: gyri and dark gray: sulci). Color scale indicates
790 classification accuracy (Acc) percentages.

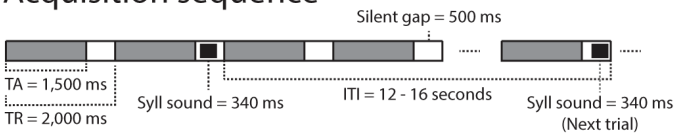
A Stimuli matrix

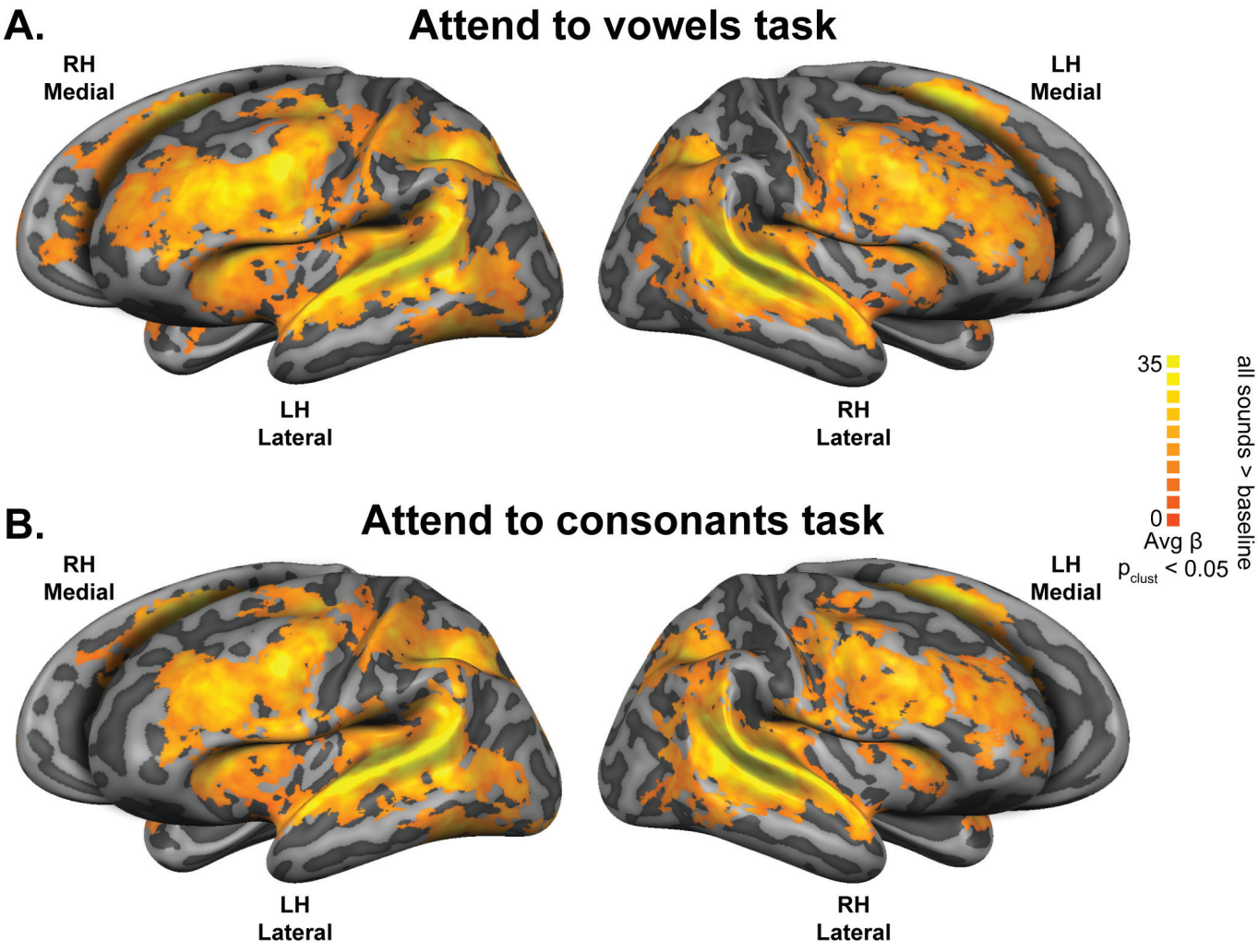


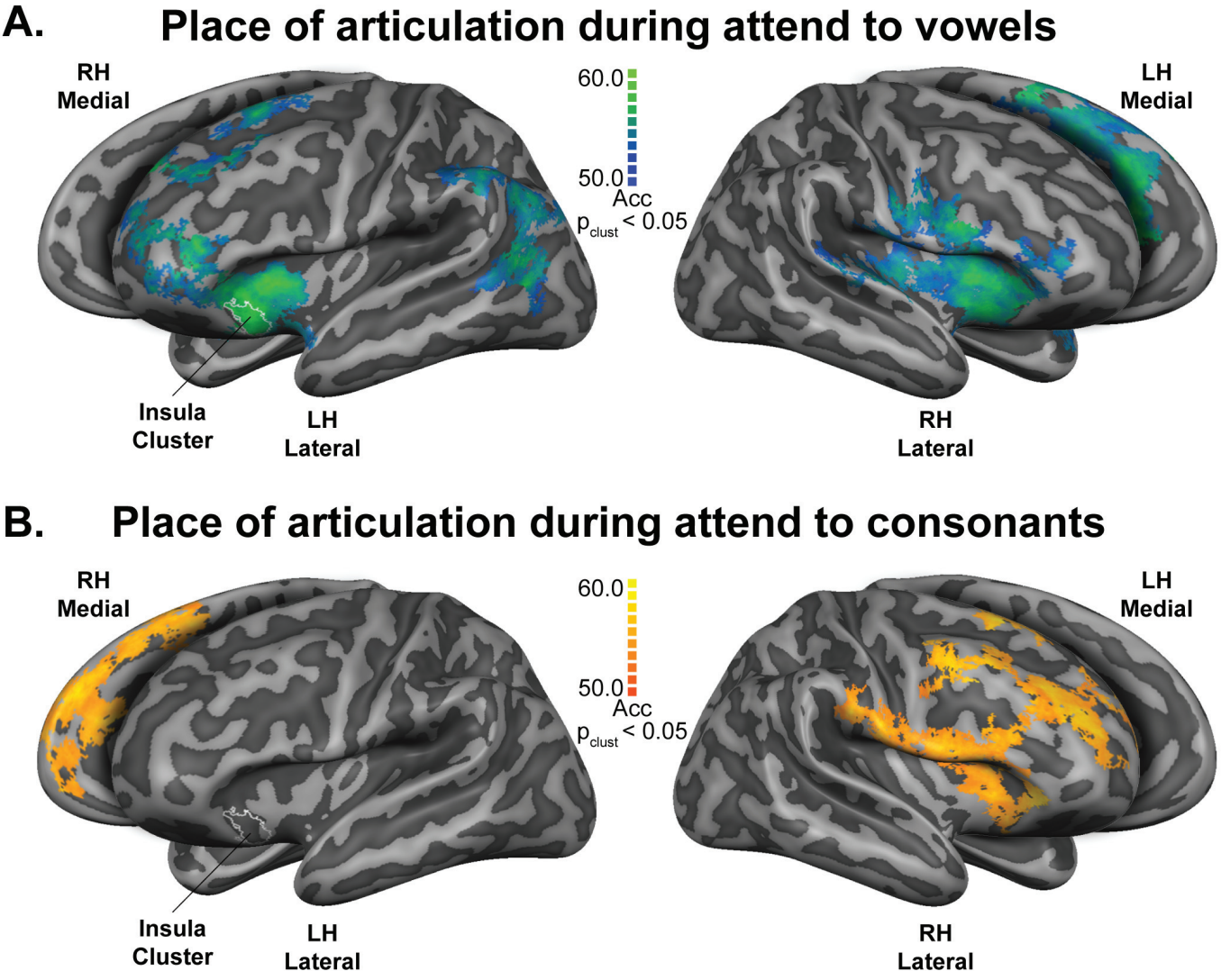
B Experiment structure

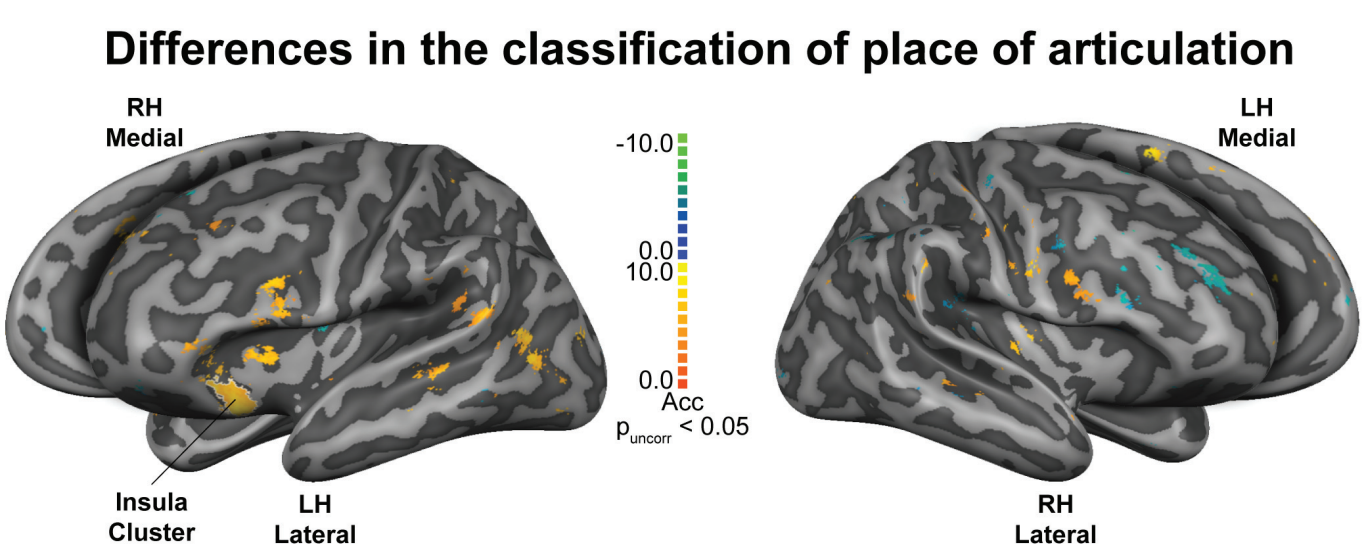


C Acquisition sequence









Place of articulation pooling all trials from both tasks