

Decoding articulatory features from fMRI responses in Dorsal speech regions

Citation for published version (APA):

Correia, J. M., Jansma, B. M., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in Dorsal speech regions. *Journal of Neuroscience*, 35(45), 15015-15025.
<https://doi.org/10.1523/jneurosci.0977-15.2015>

Document status and date:

Published: 01/01/2015

DOI:

[10.1523/jneurosci.0977-15.2015](https://doi.org/10.1523/jneurosci.0977-15.2015)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions

Joao M. Correia, Bernadette M.B. Jansma, and Milene Bonte

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, and Maastricht Brain Imaging Center, 6229 EV Maastricht, The Netherlands

The brain's circuitry for perceiving and producing speech may show a notable level of overlap that is crucial for normal development and behavior. The extent to which sensorimotor integration plays a role in speech perception remains highly controversial, however. Methodological constraints related to experimental designs and analysis methods have so far prevented the disentanglement of neural responses to acoustic versus articulatory speech features. Using a passive listening paradigm and multivariate decoding of single-trial fMRI responses to spoken syllables, we investigated brain-based generalization of articulatory features (place and manner of articulation, and voicing) beyond their acoustic (surface) form in adult human listeners. For example, we trained a classifier to discriminate place of articulation within stop syllables (e.g., /pa/ vs /ta/) and tested whether this training generalizes to fricatives (e.g., /fa/ vs /sa/). This novel approach revealed generalization of place and manner of articulation at multiple cortical levels within the dorsal auditory pathway, including auditory, sensorimotor, motor, and somatosensory regions, suggesting the representation of sensorimotor information. Additionally, generalization of voicing included the right anterior superior temporal sulcus associated with the perception of human voices as well as somatosensory regions bilaterally. Our findings highlight the close connection between brain systems for speech perception and production, and in particular, indicate the availability of articulatory codes during passive speech perception.

Key words: articulatory gestures; auditory cortex; fMRI; MVPA; sensorimotor; speech perception

Significance Statement

Sensorimotor integration is central to verbal communication and provides a link between auditory signals of speech perception and motor programs of speech production. It remains highly controversial, however, to what extent the brain's speech perception system actively uses articulatory (motor), in addition to acoustic/phonetic, representations. In this study, we examine the role of articulatory representations during passive listening using carefully controlled stimuli (spoken syllables) in combination with multivariate fMRI decoding. Our approach enabled us to disentangle brain responses to acoustic and articulatory speech properties. In particular, it revealed articulatory-specific brain responses of speech at multiple cortical levels, including auditory, sensorimotor, and motor regions, suggesting the representation of sensorimotor information during passive speech perception.

Introduction

Speech perception and production are closely linked during verbal communication in everyday life. Correspondingly, the neural processes responsible for both faculties are inherently connected (Hickok and Poeppel, 2007; Glasser and Rilling, 2008), with

sensorimotor integration subserving transformations between acoustic (perceptual) and articulatory (motoric) representations (Hickok et al., 2011). Although sensorimotor integration mediates motor speech development and articulatory control during speech production (Guenther and Vladusich, 2012), its role in speech perception is less established. Indeed, it remains unknown whether articulatory speech representations play an active role in speech perception and/or whether this is dependent on task-specific sensorimotor goals. Tasks explicitly requiring sensory-to-motor control, such as speech repetition (Caplan and Waters, 1995; Hickok et al., 2009), humming (Hickok et al., 2003), and verbal rehearsal in working memory (Baddeley et al., 1998; Jacquemot and Scott, 2006; Buchsbaum et al., 2011), activate the dorsal auditory pathway, including sensorimotor regions at the border of the posterior temporal and parietal lobes, the sylvian-

Received March 7, 2015; revised Sept. 9, 2015; accepted Oct. 5, 2015.

Author contributions: J.M.C., B.M.B.J., and M.B. designed research; J.M.C. performed research; J.M.C. contributed unpublished reagents/analytic tools; J.M.C. analyzed data; J.M.C., B.M.B.J., and M.B. wrote the paper.

This work was supported by European Union Marie Curie Initial Training Network Grant PITN-GA-2009-238593. We thank Elia Formisano for thorough discussions and advice on the study design and analysis and Matt Davis for valuable discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Joao M. Correia, Maastricht University, Oxfordlaan 55, 6229 EV Maastricht, The Netherlands. E-mail: joao.correia@maastrichtuniversity.nl.

DOI:10.1523/JNEUROSCI.0977-15.2015

Copyright © 2015 the authors 0270-6474/15/3515015-11\$15.00/0

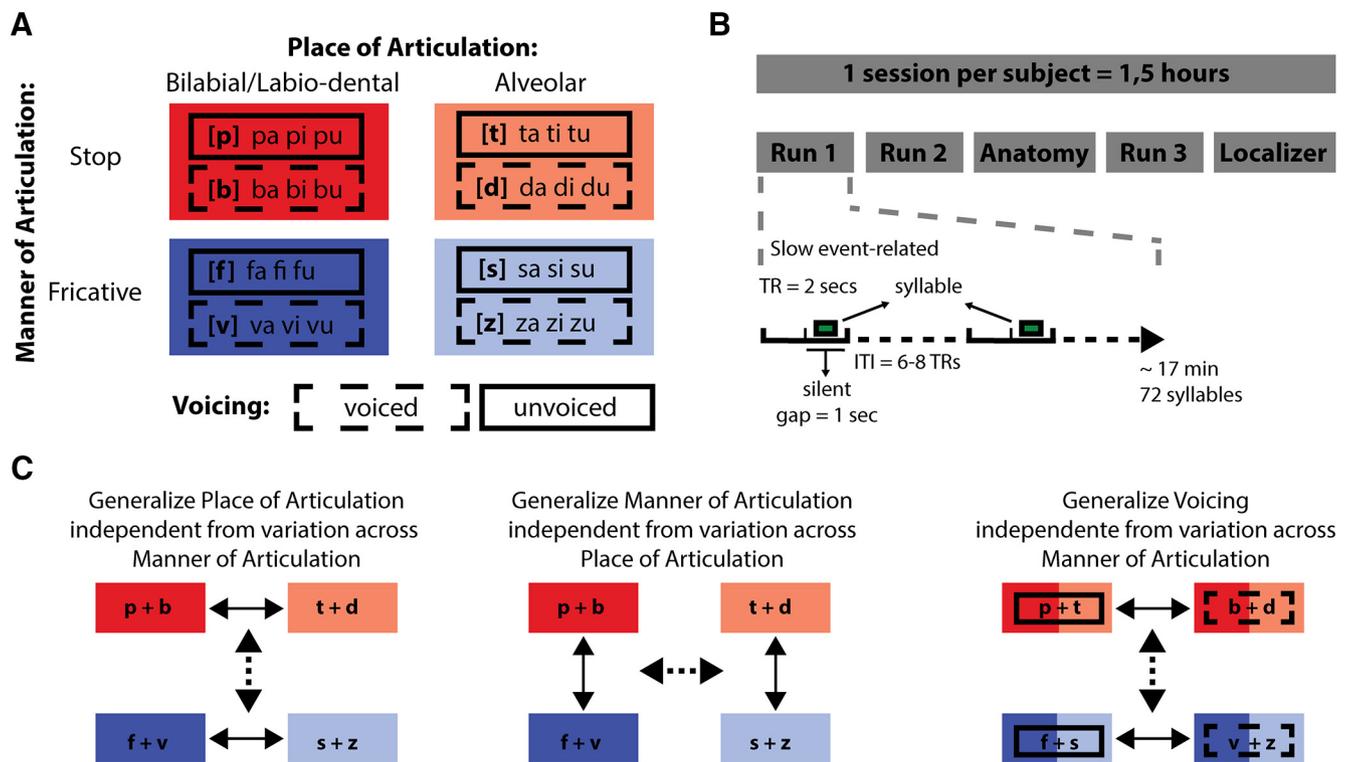


Figure 1. Experimental paradigm and analysis. **A**, The spoken syllable stimuli and their articulatory features. Stimuli were selected according to place of articulation (bilabial/labio-dental and alveolar), manner of articulation (stop and fricative), and voicing (voiced and unvoiced). **B**, The experimental session included three functional runs during which syllables were presented in a slow event-related fashion (ITI = 12–16 s) during a silent gap (1 s) between consecutive volume acquisitions. **C**, fMRI decoding analysis based on three generalization strategies: generalization of place of articulation across variation of manner of articulation; generalization of manner of articulation across variation of place of articulation; and generalization of voicing across variation of manner of articulation. Both generalization directions were tested. In all generalization analyses, all the stimuli were used.

parieto-temporal region and supramarginal gyrus (SMG) (Hickok and Poeppel, 2007). Also, in experimental paradigms that do not explicitly require sensorimotor integration, the perception of speech may involve a coactivation of motor speech regions (Zatorre et al., 1992; Wilson et al., 2004). These activations may follow a topographic organization, such as when listening to syllables involving the lips (e.g., /ba/) versus the tongue (e.g., /da/) (Pulvermüller et al., 2006), and may be selectively disrupted by transcranial magnetic stimulation (D'Ausilio et al., 2009). Whether this coinvolvement of motor areas in speech perception reflects an epiphenomenal effect due to an interconnected network for speech and language (Hickok, 2009), a compensatory effect invoked in case of noisy and/or ambiguous speech signals (Hervais-Adelman et al., 2012; Du et al., 2014), or neural computations used for an articulatory-based segmentation of speech input in everyday life situations remains unknown (Meister et al., 2007; Pulvermüller and Fadiga, 2010).

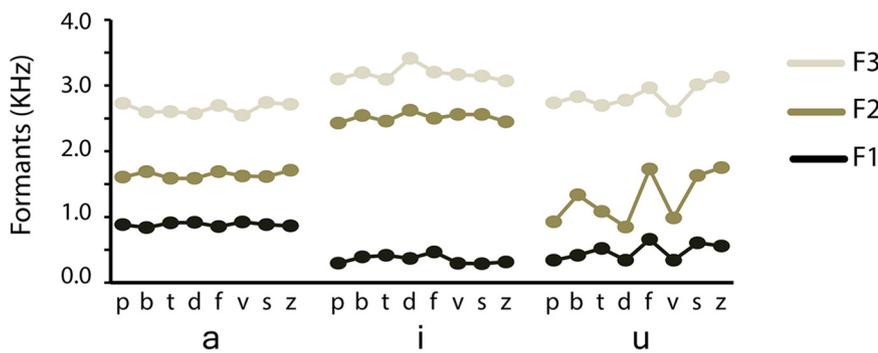
Beyond regional modulations of averaged activity across different experimental conditions, fMRI in combination with multivoxel pattern analysis (MVPA) allows investigating the acoustic and/or articulatory representation of individual speech sounds. This approach has been successful in demonstrating auditory cortical representations of speech (Formisano et al., 2008; Kilian-Hütten et al., 2011; Lee et al., 2012; Bonte et al., 2014; Arsenault and Buchsbaum, 2015; Evans and Davis, 2015). Crucially, MVPA enables isolating neural representations of stimulus classes from variation across other stimulus dimensions, such as the representation of vowels independent of acoustic variation across speakers' pronunciations (Formisano et al., 2008) or the representation of semantic concepts independent of the input language in bilingual listeners (Correia et al., 2014).

In this high-resolution fMRI study, we used a passive listening paradigm and an MVPA-based generalization approach to examine neural representations of articulatory features during speech perception with minimal sensorimotor demands. A balanced set of spoken syllables and MVPA generalization using a surface-based searchlight procedure (Kriegeskorte et al., 2006; Chen et al., 2011) allowed unraveling these representations in distinct auditory, sensorimotor, and motor regions. Stimuli consisted of 24 consonant-vowel syllables constructed from 8 consonants (/b/, /d/, /f/, /p/, /s/, /t/, /v/, and /z/) and 3 vowels (/a/, /i/, and /u/), forming two features for each of three articulatory dimensions: place and manner of articulation and voicing (Fig. 1A). A slow event-related design with an intertrial interval of 12–16 s assured a maximally independent single-trial fMRI acquisition. Our MVPA generalization approach consisted of training, for example, a classifier to discriminate between two places of articulation (e.g., /pa/ vs /ta/) for stop consonants and testing whether this training generalizes to fricatives (e.g., /fa/ vs /sa/), thereby decoding fMRI responses to speech gestures beyond individual stimulus characteristics specific to, for example, abrupt sounds such as stop consonants or sounds characterized by a noise component such as fricatives. This decoding procedure was performed for (1) place of articulation across manner of articulation, (2) manner of articulation across place of articulation, and (3) voicing across manner of articulation.

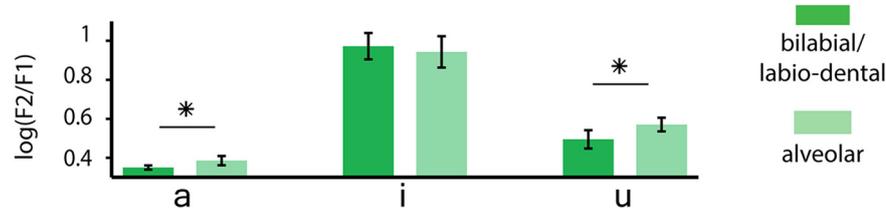
Materials and Methods

Participants. Ten Dutch-speaking participants (5 males, 5 females; mean \pm SD age, 28.2 \pm 2.35 years; 1 left handed) took part in the study. All participants were undergraduate or postgraduate students of Maastricht University, reported normal hearing abilities, and were neurolog-

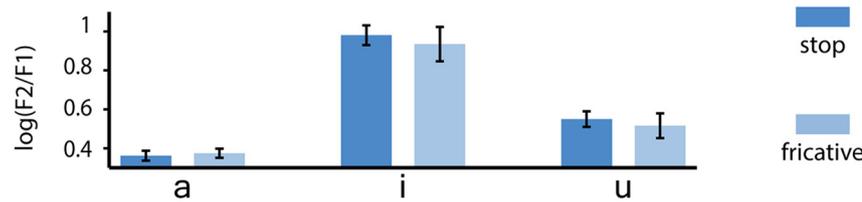
A - F1, F2 and F3 per syllable



B - Place of articulation



C - Manner of articulation



D - Voicing

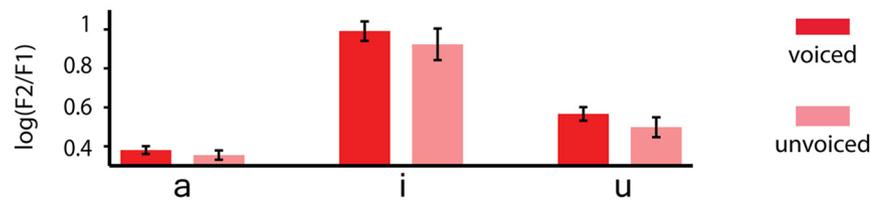


Figure 2. Acoustic properties of the vowels in coarticulatory context of the preceding consonant. **A**, Formants F1, F2, and F3 for each syllable, extracted from a 100 ms window centered at the mid-point of the vowel segments. **B–D**, Ratio $\log(F2/F1)$ separated by vowels and grouped according to the articulatory dimensions used in our MVPA generalization analysis: (**B**) place of articulation, (**C**) manner of articulation, and (**D**) voicing. * $p < 0.05$.

ically healthy. The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht, Maastricht, The Netherlands.

Stimuli. Stimuli consisted of 24 consonant-vowel (CV) syllables pronounced by 3 female Dutch speakers, generating a total of 72 sounds. The syllables were constructed based on all possible CV combinations of 8 consonants (/b/, /d/, /f/, /p/, /s/, /t/, /v/, and /z/) and 3 vowels (/a/, /i/, and /u/). The 8 consonants were selected to cover two articulatory features per articulatory dimension (Fig. 1A): for place of articulation, bilabial/labio-dental (/b/, /p/, /f/, /v/) and alveolar (/t/, /d/, /s/, /z/); for manner of articulation, stop (/p/, /b/, /t/, /d/) and fricative (/f/, /v/, /s/, /z/); and for voicing, unvoiced (/p/, /t/, /f/, /s/) and voiced (/b/, /d/, /v/, /z/). The different vowels and the three speakers introduced acoustic variability. Stimuli were recorded in a soundproof chamber at a sampling rate of 44.1 kHz (16-bit resolution). Postprocessing of the recorded stimuli was performed in PRAAT software (Boersma and Weenink, 2001) and included

bandpass filtering (80–10,500 Hz), manual removal of acoustic transients (clicks), length equalization, removal of sharp onsets and offsets using 30 ms ramp envelopes, and amplitude equalization (average RMS). Stimulus length was equated to 340 ms using PSOLA (pitch synchronous overlap and add) with 75–400 Hz as extrema of the F0 contour. Length changes were small (mean \pm SD, 61 \pm 47 ms), and subjects reported that the stimuli were unambiguously comprehended during a stimuli familiarization phase before the experiment. We further checked our stimuli for possible alterations in F0 after length equalization and found no significant changes of maximum F0 ($p = 0.69$) and minimum F0 ($p = 0.76$) with respect to the original recordings.

Acoustic characteristics of the 24 syllables were determined using PRAAT software and included the first three spectrogram formants (F1, F2, and F3) extracted from the 100 ms window centered at the midpoint of the vowel segment of each syllable (Fig. 2A). Because place of articulation of consonants is known to influence F2/F1 values of subsequent vowels in CV syllables due to coarticulation (Rietveld and van Heuven, 2001; Ladefoged and Johnson, 2010; Bouchard and Chang, 2014), we additionally calculated the logarithmic ratio of F2/F1 for the vowels in each of the syllables and assessed statistical differences between articulatory features (Fig. 2B–D). As expected, place of articulation led to significant $\log(F2/F1)$ differences for the vowel ‘u’ and ‘a’ ($p < 0.05$), with smaller $\log(F2/F1)$ values for vowels preceded by bilabial/labio-dental than alveolar consonants. No significant $\log(F2/F1)$ differences were found for the vowel ‘i’ or for any of the vowels when syllables were grouped along manner of articulation or voicing. Importantly, together with pronunciations from three different speakers, we used three different vowels to increase acoustic variability and to weaken stimulus-specific coarticulation effects in the analyses.

Experimental procedures. The main experiment was divided into three slow event-related runs (Fig. 1B). The runs consisted of randomly presenting each of the 72 sounds once, separated by an intertrial interval (ITI) of 12–16 s (corresponding to 6–8 TRs) while participants were asked to attend to the spoken syllables. Stimulus presentation was pseudorandomized such that consecutive presentations of the same syllables were avoided. Before starting the measurements, examples of the syllables were presented binaurally (using MR-compatible in-ear headphones; Sensimetrics, model S14; www.sens.com) at the same comfortable intensity level. This level was then individually adjusted according to the indications provided by each participant to equalize their perceived loudness. During scanning, stimuli were presented in silent periods (1 s) between two acquisition volumes. Participants were asked to fixate at a gray fixation cross against a black background to keep the visual stimulation constant during the entire duration of a run. Run transitions were marked with written instructions. Although monitoring of possible subvocal/unconscious rehearsal accompanying the perception of the spoken syllables was not conducted, none of the participants reported the use of subvocal rehearsal strategies.

fMRI acquisition. Functional and anatomical image acquisition was performed on a Siemens TRIO 3 tesla scanner (Scannexus) at the Maas-

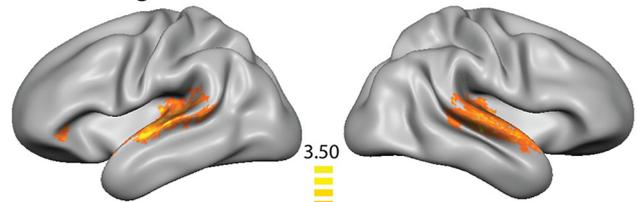
tricht Brain Imaging Center. Functional runs used in the main experiment were collected per subject with a spatial resolution of 2 mm isotropic using a standard echo-planar imaging sequence [repetition time (TR) = 2.0 s; acquisition time (TA) = 1.0 s; field of view = 192 × 192 mm; matrix size = 64 × 64; echo time (TE) = 30 ms; multiband factor 2]. Each volume consisted of 25 slices aligned and centered along the Sylvian fissures of the participants. The duration difference between the TA and TR introduced a silent period used for the presentation of the auditory stimuli. High-resolution (voxel size 1 mm³ isotropic) anatomical images covering the whole brain were acquired after the second functional run using a T1-weighted 3D Alzheimer's Disease Neuroimaging Initiative sequence (TR = 2050 ms; TE = 2.6 ms; 192 sagittal slices).

Two additional localizer runs presented at the end of the experimental session were used to identify fMRI activations related to listening and repeating the spoken syllables and to guide the multivariate decoding analysis conducted in the main experiment. For the first localizer run, participants were instructed to attentively listen to the spoken syllables. For the second localizer run, participants were instructed to listen and repeat the spoken syllables. This run was presented at the end of the scanning session to prevent priming for vocalizations during the main experiment. Both localizer runs consisted of 9 blocks of 8 syllables, with one syllable presented per TR. The blocks were separated by an ITI of 12.5–17.5 s (5–7 TRs). The scanning parameters were the same as used in the main experiment, with the exception of a longer TR (2.5 s) that assured that participants were able to listen and repeat the syllables in the absence of scanner noise (silent period = 1.5 s). Figure 3, A and B, shows the overall BOLD activation evoked during the localizer runs. Listening to the spoken syllables elicited activation in the superior temporal lobe in both hemispheres (Fig. 3A), as well as in inferior frontal cortex/anterior insula in the left hemisphere. Repeating the spoken syllables additionally evoked activation in premotor (anterior inferior precentral gyrus and posterior inferior frontal gyrus), motor (precentral gyrus), and somatosensory (postcentral gyrus) regions (Fig. 3B). BOLD activation in these regions was statistically assessed using random-effects GLM statistics ($p < 0.05$) and corrected for multiple comparisons using cluster size threshold correction ($\alpha = 5\%$). We defined an ROI per hemisphere that was used for the decoding analysis of the main experiment (Fig. 3C). The ROI included parts of the temporal lobes, inferior frontal cortices and parietal lobes, which are typically activated in speech perception and production tasks. By comparing the ROI with the activity obtained with the localizers, we made sure that areas activated during the perception and repetition of spoken syllables were all included.

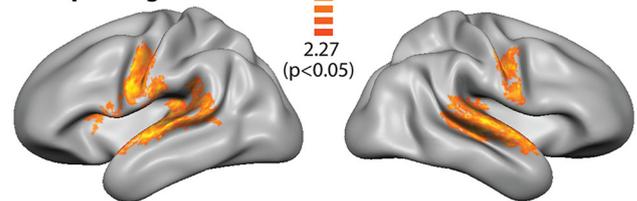
fMRI data preprocessing. fMRI data were preprocessed and analyzed using Brain Voyager QX version 2.8 (Brain Innovation) and custom-made MATLAB (The MathWorks) routines. Functional data were 3D motion-corrected (trilinear sinc interpolation), corrected for slice scan time differences, and temporally filtered by removing frequency components of ≤ 5 cycles per time course (Goebel et al., 2006). According to the standard analysis scheme in Brain Voyager QX (Goebel et al., 2006), anatomical data were corrected for intensity inhomogeneity and transformed into Talairach space (Talairach and Tournoux, 1988). Individual cortical surfaces were reconstructed from gray-white matter segmentations of the anatomical acquisitions and aligned using a moving target-group average approach based on curvature information (cortex-based alignment) to obtain an anatomically aligned group-averaged 3D surface representation (Goebel et al., 2006; Frost and Goebel, 2012). Functional data were projected to the individual cortical surfaces, creating surface-based time courses. All statistical analyses were then conducted on the group averaged surface making use of cortex-based alignment.

MVPA classification (generalization of articulatory features). To investigate the local representation of spoken syllables based on their articulatory features, we used multivariate classification in combination with a moving searchlight procedure that selected cortical vertices based on their spatial (geodesic) proximity. The particular aspect pursued by the MVPA was to decode articulatory features of the syllables beyond their phoneme specific acoustic signatures (Fig. 1C). Hence, we used a classification strategy based on the generalization of articulatory features across different types of phonemes. For example, we trained a classifier to decode place of articulation features (bilabial/labio-dental vs alveolar

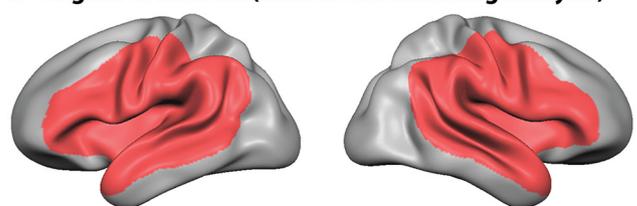
A - Listening localizer



B - Repeating localizer



C - Region of interest (used in the decoding analysis)



D - Searchlight radii

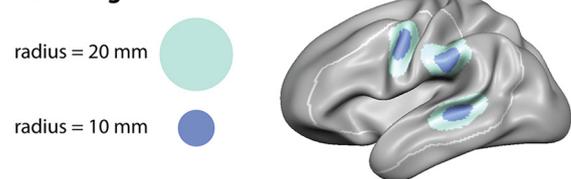


Figure 3. Group-level cortical activity maps (based on random effects GLM analysis, $p < 0.05$) of the (A) passive listening and (B) verbal repetition localizer runs. C, A group ROI per hemisphere was anatomically defined to include speech-related perisylvian regions and the activation elicited by the localizers. D, Examples of cortical spread used in the surface-based searchlight analyses. Light blue represents radius = 10 mm. Dark blue represents radius = 20 mm.

from stop syllables ([/b/ and /p/] vs [/t/ and /d/]) and tested whether this learning is transferable to fricative syllables ([/f/ and /v/] vs [/s/ and /z/]), thus decoding place of articulation features across phonemes differing in manner of articulation. In total, we performed such a generalization strategy to investigate the neural representation of place of articulation (bilabial/labio-dental vs alveolar) across manner of articulation (stop and fricatives); manner of articulation (stop vs fricatives) across place of articulation (bilabial/labio-dental and alveolar); and voicing (voiced vs unvoiced) across manner of articulation (stop and fricatives). Additional methodological steps encompassing the construction of the fMRI feature space (fMRI feature extraction and surface-based searchlight procedure) as well as the computational strategy to validate (cross-validation) and display (generalization maps) the classification results are described below in detail.

fMRI feature extraction. Before classification, BOLD responses of each fMRI trial and each cortical vertex were estimated by fitting a hemodynamic response function using a GLM. To account for the temporal variability of single-trial BOLD responses, multiple hemodynamic response function fittings were produced by shifting their onset time (lag) with respect to the stimulus event time (number of lags = 21, interval between consecutive lags = 0.1 s) (Ley et al., 2012; De Martino et al., 2008). At each trial, the GLM coefficient β resulting from the best fitting hemodynamic response function across lags in the whole brain was used to construct an fMRI feature space composed by the number of trials by

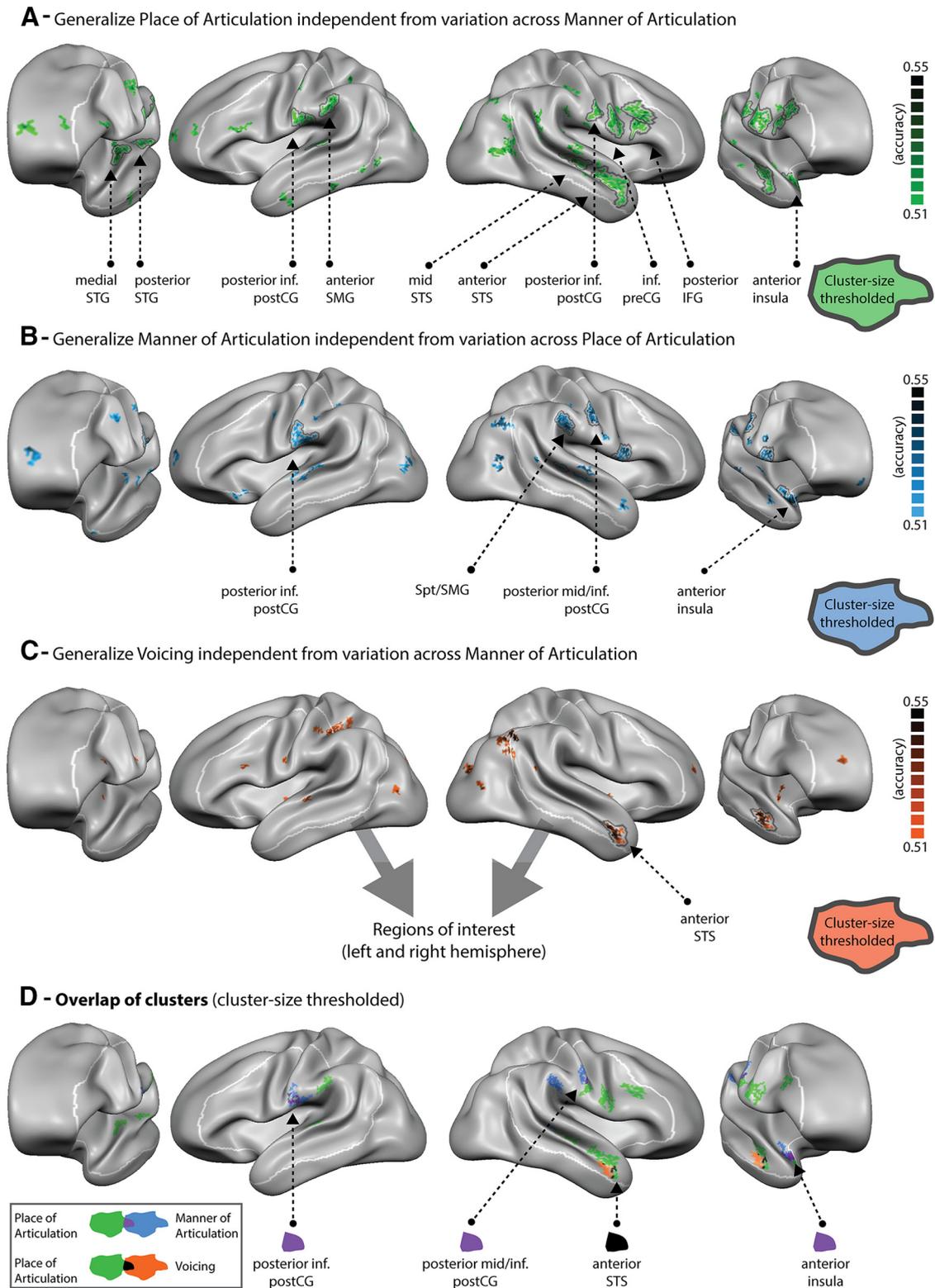


Figure 4. Maps of group-averaged accuracy values for (A) place of articulation, (B) manner of articulation, and (C) voicing as well as (D) the spatial overlap of these three maps after cluster size correction. All maps are projected onto the group-aligned cortical surface with a minimum cluster size of 3 mm² and statistically thresholded ($p < 0.05$). Group-level maps were statistically assessed using random-effects statistics (exact permutation tests = 1022) and corrected for multiple comparisons using cluster size correction (1000 iterations). Black line indicates regions surviving cluster size correction threshold (> 18 mm²). White line indicates preselected left and right hemisphere ROIs. Purple colored vertices represent the overlap between regions generalizing place of articulation and manner of articulation. Black colored vertices represent the overlap between regions generalizing place of articulation and voicing. No overlap was found between manner of articulation and voicing.

the number of cortical vertices, which was thereafter used in the multivariate decoding.

Surface-based searchlight procedure (selecting cortical vertices for classification). To avoid degraded performances of the classification algorithm attributable to the high dimensionality of the feature space (model overfitting; for a description, see Norman et al., 2006), a reduction of the number of fMRI features is usually performed. The moving searchlight approach (Kriegeskorte et al., 2006) restricts fMRI features using focal selections centered at all voxels within spherical patches of the gray matter volume. Here, we used a searchlight procedure on the gray-white matter segmentation surface (Chen et al., 2011), which selected cortical vertices for decoding based on their spatial (geodesic) distance within circular surface patches with a radius of 10 mm (Fig. 3D). The surface-based searchlight procedure reduces the concurrent inclusion of voxels across different gyri that are geodesically distant but nearby in 3D volume space, and has been shown reliable for fMRI MVPA (Chen et al., 2011; Oosterhof et al., 2011). Crucially, the surface-based searchlight procedure assures an independent analysis of superior temporal and ventral frontal cortical regions that may be involved in the articulatory representation of speech. The primary searchlight analysis was conducted in the predefined ROI comprising perisylvian speech and language regions (Fig. 3C). An additional exploratory analysis in the remaining cortical surface covered by the functional scans was also conducted. Furthermore, next to the searchlight analysis using a radius similar to that used in previous speech decoding studies (Lee et al., 2012; Correia et al., 2014; Evans and Davis, 2015), we performed a further analysis using a larger searchlight radius of 20 mm (Fig. 3D). Different radius sizes in the searchlight method may exploit different spatial spreads of fMRI response patterns (Nestor et al., 2011; Etzel et al., 2013).

Cross-validation. Cross-validation was based on the generalization of articulatory features of syllables independent of acoustic variation across other articulatory features. For each classification strategy, two cross-validation folds were created and included generalization in one and the opposite direction (e.g., generalization of place of articulation from stop to fricatives and from fricatives to stop syllables). Cross-validation based on generalization strategies is attractive because it enables detecting activation patterns resistant to variation across other stimuli dimensions (Formisano et al., 2008; Buchweitz et al., 2012; Correia et al., 2014). As we aimed to maximize the acoustic variance of our decoding scheme, generalization of place of articulation and voicing were both calculated across manner of articulation, the dimension that is acoustically most distinguishable (Rietveld and van Heuven, 2001; Ladefoged and Johnson, 2010). Generalization of manner of articulation was performed across place of articulation.

Generalization maps. At the end of the searchlight decoding procedure, individual averaged accuracy maps for place of articulation, manner of articulation, and voicing were constructed, projected onto the group-averaged cortical surface, and anatomically aligned using cortex-based alignment. To assess group-averaged statistical significance of cortical vertices (chance level is 50%), exact permutation tests were used ($n = 1022$). The resulting statistical maps were then corrected for multiple comparisons by applying a cluster size threshold with a false-positive rate

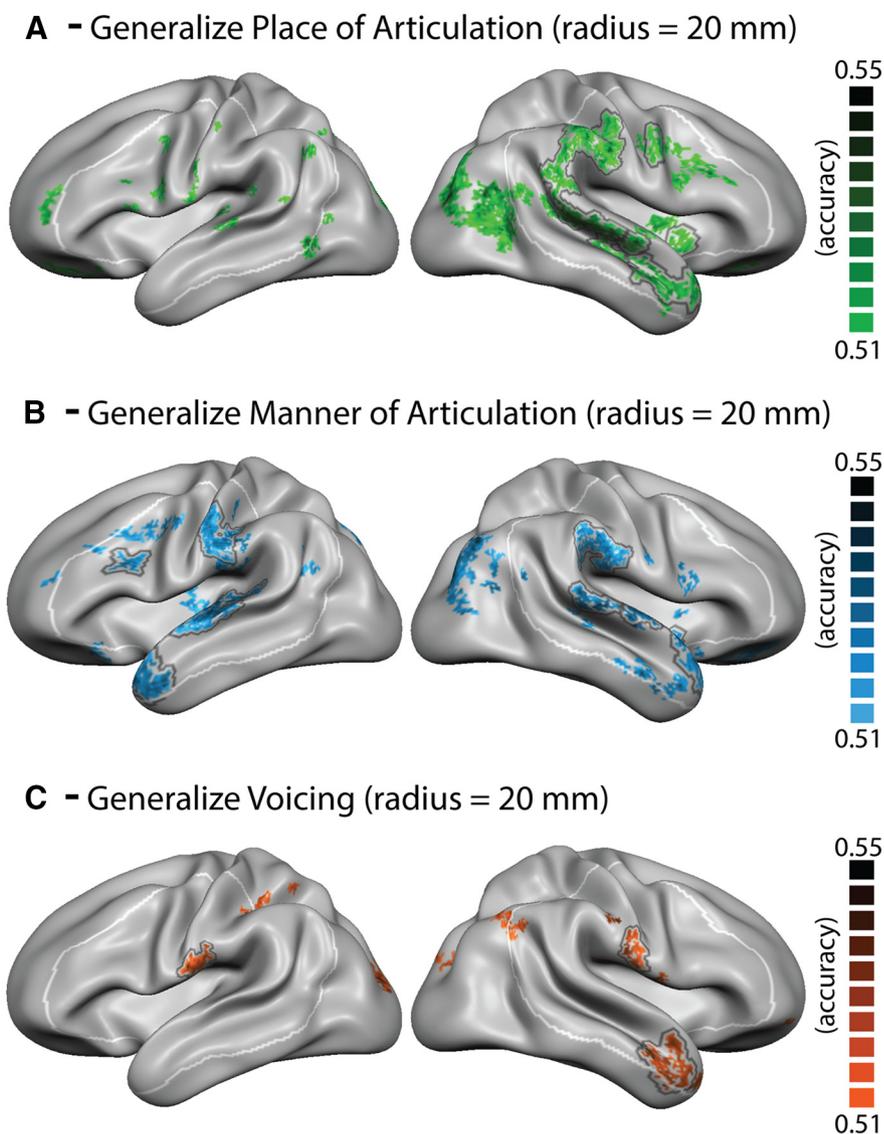


Figure 5. Maps of group-averaged accuracy values of the searchlight analysis using a 20 mm radius. **A**, Place of articulation. **B**, Manner of articulation. **C**, Voicing. All maps are projected onto the group-aligned cortical surface with a minimum cluster size of 3 mm^2 and statistically thresholded ($p < 0.05$). Group-level maps were statistically assessed using random-effects statistics (exact permutation tests = 1022) and corrected for multiple comparisons using cluster size correction (1000 iterations). Black line indicates regions surviving cluster size threshold ($>41 \text{ mm}^2$). White line indicates preselected left and right hemisphere ROIs.

($\alpha = 5\%$) after setting an initial vertex-level threshold ($p < 0.05$, uncorrected) and submitting the maps to a correction criterion based on the estimate of the spatial smoothness of the map (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006).

Results

To investigate the neural representation of articulatory features during speech perception, we used a classification strategy that relied on generalizing the discriminability of three articulatory features across different syllable pairs. That is, using a moving searchlight procedure, we tested which cortical regions allow the generalization of (1) place of articulation across two distinct manners of articulation (Figs. 4A, 5A), (2) manner of articulation across two distinct places of articulation (Figs. 4B, 5B), and (3) voicing across two distinct manners of articulation (Figs. 4C, 5C). For each type of generalization analysis, half of the available trials was used to learn to discriminate the neural responses evoked by the articulatory features of interest and the other half was used to

test the generalization of this learning. The same stimuli were hence included in all generalization analyses in a counterbalanced manner.

Figure 4A–C depicts the generalization accuracy maps for the three types of decoding obtained from the primary analysis, statistically masked ($p < 0.05$) with black circled clusters indicating regions that survived cluster size multiple comparisons correction. The generalization maps revealed successful decoding of each of the three articulatory features within distinct but partially overlapping regions of the brain's speech perception network. We observed generalization foci within both the left and right hemispheres, suggesting the participation of bilateral language and speech regions in the representation of spoken syllables based on their articulatory/motoric properties. Cortical regions within the ROI (Fig. 3C) enabling the generalization of place and manner of articulation were most widely distributed, including both superior temporal, premotor, motor and somatosensory regions bilaterally, as well as sensorimotor areas at the border between the parietal and temporal lobes in the left hemisphere in case of place of articulation and in the right hemisphere in case of manner of articulation. Specific regions leading to significant generalization of place of articulation included localized clusters in the left medial and posterior superior temporal gyrus (STG), left posterior inferior post-central gyrus (CG) and left anterior SMG, as well as right middle and anterior superior temporal sulcus (STS), right posterior inferior post-CG, inferior pre-CG, and right posterior inferior frontal gyrus (IFG). Generalization of manner of articulation was significantly possible based on clusters within the left posterior inferior post-CG, right mid STS, right posterior mid/inferior post-CG, inferior anterior pre-CG/posterior IFG, right anterior SMG, and right anterior insula. In contrast to the contribution of multiple, distributed brain activity clusters to the representation of place and manner of articulation, generalization of voicing across different syllables was more confined, including specifically the right anterior STS. Finally, a visual comparison of the three types of generalization maps (Fig. 4D) demonstrates spatially overlapping clusters of significant generalization. Overlap between place and manner of articulation was observed within the inferior post-CG bilaterally as well as in the right anterior precentral gyrus/posterior IFG and the right anterior insula. Overlap between place of articulation and voicing was observed within the right anterior STS.

An exploratory analysis beyond our speech-related perisylvian ROI showed additional clusters with the capacity to generalize articulatory properties (Fig. 4; Table 1). Specifically, place of articulation was significantly decodable in the right posterior middle temporal gyrus, left cuneus and precuneus, as well as the caudate and cerebellum bilaterally. Clusters allowing the generalization of manner of articulation were found in the right inferior angular gyrus, as well as in the right parahippocampal gyrus and cerebellum. Finally, additional clusters for voicing included the left intraparietal sulcus, the right inferior angular gyrus as well as the left anterior cingulate, right cuneus and precuneus, and the right parahippocampal gyrus.

To investigate the contribution of cortical representations with a larger spatial extent, we performed an additional analysis using a searchlight radius of 20 mm (Fig. 5; Table 2). The larger searchlight radius yielded broader clusters, as expected, and the overall pattern of results was comparable with that obtained with the 10 mm searchlight radius (Fig. 4). It also yielded significant effects in additional clusters within the ROI. In particular, generalization of manner of articulation led to significant clusters along the superior temporal lobe bilaterally and in the left inferior

Table 1. Whole-brain clusters for the searchlight analysis using a radius of 10 mm^a

Brain region	Talairach			<i>p</i> value (peak)	Cluster size (mm ²)
	<i>x</i>	<i>y</i>	<i>z</i>		
Place of articulation					
LH Media STG	−48	−11	6	0.002930	44
LH Posterior STG	−54	−24	9	0.002930	20
LH Posterior inferior post-CG	−52	−18	20	0.000977	18
LH Anterior SMG	−58	27	23	0.002930	24
LH Precuneus	−11	−61	50	0.011719	42
LH Subcallosal gyrus	−14	15	−12	0.001953	48
LH Caudate	−1	4	16	0.000977	87
LH Caudate	−1	18	4	0.001953	25
LH Cuneus	−8	−77	27	0.000977	20
RH Anterior insula	27	2	−17	0.000977	66
RH IFG (superior)	39	13	31	0.042969	33
RH IFG (inferior)	47	12	24	0.032227	18
RH Anterior STS	48	−1	−14	0.005859	38
RH Inferior pre-CG	48	−1	27	0.009766	34
RH Posterior inferior post-CG	61	−13	24	0.042969	28
RH Mid-STS	50	−20	1	0.004883	21
RH Caudate	1	14	7	0.022461	50
RH Cerebellum	20	−29	−15	0.032227	46
RH Cerebellum	18	−43	−7	0.009766	29
RH Cerebellum	14	−75	33	0.011719	22
RH Posterior middle temporal gyrus	43	−55	4	0.006836	36
Manner of articulation					
LH Posterior inferior post-CG	−53	−18	22	0.000977	45
RH Anterior insula	27	2	−17	0.000977	62
RH IFG	52	8	13	0.046641	18
RH Posterior mid/inferior post-CG	56	−12	36	0.001953	19
RH SMG	51	−27	27	0.002930	26
RH Parahippocampal gyrus	22	−6	−24	0.000977	26
RH AG	45	−60	30	0.000977	23
RH Cerebellum	27	−69	−13	0.033203	21
Voicing					
LH IPS	−36	−37	39	0.000977	66
LH Anterior Cingulate	−11	37	18	0.002930	20
RH Anterior STS	47	5	−18	0.031250	31
RH AG	44	−59	30	0.005859	18
RH Parahippocampal gyrus	26	−43	−9	0.018555	18
RH Precuneus	12	−47	29	0.001953	19
RH Cuneus	23	−80	16	0.026367	21

^aTalairach coordinates (center of gravity) in the left and right hemispheres, peak *p* value, and size of cluster.

frontal gyrus. Significant decoding of place of articulation was found in the right SMG, and the generalization of voicing led to additional significant clusters in the inferior post-CG bilaterally. Finally, decreases of generalization for the larger searchlight radius were found, especially within the left superior temporal lobe, posterior inferior post-CG, and anterior SMG for place of articulation.

Discussion

The present study aimed to investigate whether focal patterns of fMRI responses to speech input contain information regarding articulatory features when participants are attentively listening to spoken syllables in the absence of task demands that direct their attention to speech production or monitoring. Using high spatial resolution fMRI in combination with an MVPA generalization approach, we were able to identify specific foci of brain activity that discriminate articulatory features of spoken syllables independent of their individual acoustic variation (surface form) across other articulatory dimensions. These results provide compelling evidence for interlinked brain circuitry of speech perception and production within the dorsal speech regions, and in particular, for the availability of articulatory codes during online

Table 2. Whole-brain clusters for the searchlight analysis using a radius of 20 mm^a

Brain region	Talairach			<i>p</i> value (peak)	Cluster size (mm ²)
	<i>x</i>	<i>y</i>	<i>z</i>		
Place of articulation					
LH Caudate	−1	1	5	0.001953	246
LH IFG (anterior)	−21	34	−10	0.014648	115
LH Cuneus	−17	−90	14	0.002930	122
LH Superior frontal gyrus (medial)	−8	50	24	0.003906	32
RH Anterior STS/insula	33	13	−23	0.000977	405
RH Posterior STG/SMG	56	−26	3	0.003906	419
RH Inferior pre-CG	54	−6	32	0.018555	41
RH Precuneus	26	−75	22	0.002930	547
RH Parahippocampal gyrus	21	−24	−17	0.025391	83
RH Posterior cingulate	7	−43	6	0.012695	111
RH Anterior cingulate	4	20	20	0.033203	160
RH Inferior frontal gyrus (inferior)	23	32	−6	0.033203	63
Manner of articulation					
LH STG	−54	−12	5	0.011719	169
LH ATL	−47	12	−19	0.003906	107
LH Posterior inferior post-CG	−56	−18	27	0.013672	97
LH Inferior frontal gyrus	−43	19	21	0.010742	41
LH Anterior cingulate	−1	17	14	0.016602	48
LH Caudate	−1	1	5	0.001953	48
LH Precuneus	−14	−73	41	0.042734	44
RH Anterior STG/insula	36	17	−23	0.003906	370
RH SMG	51	−32	23	0.013672	103
RH STG	62	−18	9	0.013672	76
RH Middle frontal gyrus	26	43	−5	0.014648	139
RH Cerebellum	15	−52	−5	0.015625	66
RH Cuneus	7	−71	17	0.010742	90
RH Inferior AG/posterior MTG	40	−77	17	0.005859	141
RH Cingulate gyrus	6	−31	36	0.007812	341
Voicing					
LH Posterior inferior post-CG	−60	−15	18	0.000977	46
LH Middle occipital gyrus	−22	−89	12	0.001953	48
RH Anterior STS	47	8	−26	0.000977	100
RH Posterior inferior post-CG	60	−12	30	0.000977	44

^aTalairach coordinates (center of gravity) in the left and right hemispheres, peak *p* value, and size of cluster.

perception of spoken syllables within premotor and motor, somatosensory, auditory, and/or sensorimotor integration areas.

Our generalization analysis suggests the involvement of premotor and motor areas in the neural representation of two important articulatory features during passive speech perception: manner and place of articulation. These findings are compelling as the role of premotor and motor areas during speech perception remains controversial (Lieberman and Mattingly, 1985; Galantucci et al., 2006; Pulvermüller and Fadiga, 2010). Left hemispheric motor speech areas have been found to be involved both in the subvocal rehearsal and perception of the syllables /ba/ and /da/ (Pulvermüller et al., 2006) and of spoken words (Schomers et al., 2014). Furthermore, speech motor regions may bias the perception of ambiguous speech syllables under noisy conditions (D'Ausilio et al., 2009; Du et al., 2014) and have been suggested to be specifically important for the performance of tasks requiring subvocal rehearsal (Hickok and Poeppel, 2007; Krieger-Redwood et al., 2013). However, the involvement of (pre)motor cortex in speech perception may also reflect an epiphenomenal consequence of interconnected networks for speech perception and production (Hickok, 2009). Importantly, the observed decoding and generalization capacity of activation patterns in (pre)motor regions for place across variation in manner and manner across variation in place of articulation indicate the representation of articulatory information beyond mere activation spread also while passively listening to clear spoken syllables. Further investigations exploiting correlations of articulatory MVPA representations to behavioral measures of speech perception and their modulation by

task difficulty (Raizada and Poldrack, 2007) may permit mapping aspects related to speech intelligibility and may lead to a further understanding of the functional relevance of such articulatory representations.

Our results also show decoding of articulatory features in bilateral somatosensory areas. In particular, areas comprising the inferior posterior banks of the postcentral gyri were sensitive to the generalization of place and manner of articulation (Fig. 4), and of voicing when using a larger searchlight radius (Fig. 5). Somatosensory and motoric regions are intrinsically connected, allowing for the online control of speech gestures and proprioceptive feedback (Hickok et al., 2011; Bouchard et al., 2013). Together with feedback from auditory cortex, this somatosensory feedback may form a state feedback control system for speech production (SFC) (Houde and Nagarajan, 2011). The involvement of somatosensory areas in the representation of articulatory features during passive speech perception extends recent findings showing the involvement of these regions in the neural decoding of place and manner of articulation during speech production (Bouchard et al., 2013), and of place of articulation during an active perceptual task in English listeners (Arsenault and Buchsbaum, 2015). In particular, they may indicate automatic information transfer from auditory to somatosensory representations during speech perception (Cogan et al., 2014) similar to their integration as part of SFC systems for speech production.

Especially in the auditory cortex, it is essential to disentangle brain activity indicative of articulatory versus acoustic features. So far, methodological constraints related to experimental designs and analysis methods have often prevented this differentiation. Moreover, it is likely that multiple, different types of syllable representations are encoded in different brain systems responsible for auditory- and articulatory-based analysis (Cogan et al., 2014; Mesgarani et al., 2014; Evans and Davis, 2015). In a recent study, intracranial EEG responses to English phonemes indicated a phonetic organization in the left superior temporal cortex, especially in terms of manner of articulation and to a lesser extent also of place of articulation and voice onset time (Mesgarani et al., 2014). Our fMRI decoding findings show and confirm the expected encoding of manner and place in bilateral superior temporal cortex. Most relevantly, our generalization analysis suggests the existence of encoding of articulatory similarities across sets of acoustically different syllables. In particular, auditory (STG) response patterns distinguishing place of articulation in one set of speech sounds (e.g., stop consonants) were demonstrated to predict place of articulation in another set of speech sounds (e.g., fricatives). In connected speech, such as our natural consonant-vowel syllables, place of articulation also induces specific coarticulatory cues that may contribute to its neural representation (Bouchard and Chang, 2014). Importantly however, although our stimuli showed a smaller F2/F1 ratio for vowel /u/ and vowel /a/ due to coarticulation following consonants with a bilabial/labio-dental versus alveolar place of articulation, this coarticulatory cue was not present for vowel /i/. Thus, the success of our cross-generalization analysis in the presence of the balanced variance of pronunciations per consonant (vowels /a/, /i/ and /u/ and pronunciations from three different speakers) suggests that it is possible to study the representation of articulatory, in addition to auditory, features in auditory cortex. Finally, the finding that auditory cortical generalization of the most acoustically distinguishable feature, manner of articulation, was mainly present in the analysis using a larger searchlight radius is consistent with distributed phonetic representation of speech in these regions (Formisano et al., 2008; Bonte et al., 2014). Al-

though these findings also suggest a different spatial extent of auditory cortical representations for place versus manner features, the exact nature of the underlying representations remains to be determined in future studies.

Our cortical generalization maps not only show that it was possible to predict manner and place of articulation based on activation patterns in bilateral speech-related auditory areas (STG), but also to predict place of articulation and voicing based on patterns within the right anterior superior temporal lobe (STS). Articulation and especially voicing-related representations within the right anterior STS may relate to its involvement in the processing of human voices (Belin et al., 2000), and possibly to the proposed specialization of this area in perceiving vocal tract properties of speakers (e.g., shape and characteristics of the vocal folds). A role of this phonological feature in the processing of human voices is also compatible with the previous finding that voicing was more robustly classified than either place or manner of articulation when subjects performed an active gender discrimination task (Arsenault and Buchsbaum, 2015).

Decoding of articulatory representations that may relate to sensorimotor integration mechanisms, thus possibly involving the translation between auditory and articulatory codes, included regions within the inferior parietal lobes. Specifically, the anterior SMG was found to generalize manner of articulation in the right hemisphere and place of articulation in the left (10 mm searchlight radius) and right (20 mm searchlight radius) hemisphere. Nearby regions involving the inferior parietal lobe (Raizada and Poldrack, 2007; Moser et al., 2009; Kilian-Hütten et al., 2011) and sylvian-parietal-temporal regions (Caplan and Waters, 1995; Hickok et al., 2003, 2009; Buchsbaum et al., 2011) have been implicated in sensorimotor integration during speech perception as well as in mapping auditory targets of speech sounds before the initiation of speech production (Hickok et al., 2009; Guenther and Vladusich, 2012). Here, we show the sensitivity of SMG to represent articulatory features of spoken syllables during speech perception in the absence of an explicit and active task, such as repetition (Caplan and Waters, 1995; Hickok et al., 2009) or music humming (Hickok et al., 2003). Furthermore, our results suggest the involvement of inferior parietal lobe regions in the perception of clear speech, extending previous findings showing a significant role in the perception of ambiguous spoken syllables (Raizada and Poldrack, 2007) and the integration of ambiguous spoken syllables with lip-read speech (Kilian-Hütten et al., 2011).

Beyond regions in speech-related perisylvian cortex (ROI used), our exploratory whole-brain analysis suggests the involvement of additional clusters in parietal, occipital, and medial brain regions (Table 1). In particular, regions, such as, for example, the angular gyrus and intraparietal sulcus, are structurally connected to superior temporal speech regions via the middle longitudinal fasciculus (Seltzer and Pandya, 1984; Makris et al., 2013; Dick et al., 2014) and may be involved in the processing of vocal features (Petkov et al., 2008; Merrill et al., 2012). Their functional role in the sensorimotor representation of speech needs to be determined in future studies, for example, using different perceptual and sensorimotor tasks. Moreover, it is possible that different distributions of informative response patterns are best captured by searchlights of different radii and shapes (Nestor et al., 2011; Etzel et al., 2013). Our additional searchlight results obtained with an increased radius of 20 mm (Fig. 5) validated the results from the primary analysis using a radius of 10 mm (Fig. 4). They also

showed significant decoding accuracy in the bilateral superior temporal lobe for manner of articulation. Furthermore, decreases of generalization for the larger searchlight radius were also found, especially in the left sensorimotor and auditory regions for place of articulation. A similar pattern of results was previously reported in a study on the perception of ambiguous speech syllables, with a smaller radius allowing decoding in premotor regions and a larger radius allowing decoding in auditory regions (Lee et al., 2012). Whereas increases of generalization capability in given locations with a larger searchlight may indicate a broader spatial spread of the underlying neural representations, decreases of generalization in other locations may reflect the inability of more localized multivariate models to detect these representations. This also points to the important issue of how fMRI features are selected for MVPA. Alternative feature selection methods, such as recursive feature elimination (De Martino et al., 2008), optimize the number of features recursively starting from large ROIs. However, these methods are computationally very heavy and require a much larger number of training trials. Furthermore, the searchlight approach provides a more direct link between classification performance and cortical localization, which was a major goal in this study. Another relevant aspect for improving MVPA in general and for our generalization strategy in particular is the inclusion of additional variance in the stimulus set used for learning the multivariate model. For instance, future studies aiming to particularly target place of articulation distinctions could, in addition to the variations along manner of articulation used (stop and fricatives), include nasal consonants, such as the bilabial 'm' and alveolar 'n.'

Overall, the combination of the searchlight method with our generalization strategy was crucial to disentangle the neural representation of articulatory and acoustic differences between individual spoken syllables during passive speech perception. In particular, it allowed localizing the representation of bilabial/labio-dental versus alveolar (place), stop versus fricative (manner), and voiced versus unvoiced (voicing) articulatory features in multiple cortical regions within the dorsal auditory pathway that are relevant for speech processing and control (Hickok and Poeppel, 2007). Similar generalization strategies capable of transfer representation patterns across different stimulus classes have been adopted in MVPA studies, for example, in isolating the identity of vowels and speakers independent of acoustic variation (Formisano et al., 2008), in isolating concepts independent of language presentation in bilinguals (Buchweitz et al., 2012; Correia et al., 2014, 2015), and in isolating concepts independent of presentation modality (Shinkareva et al., 2011; Simanova et al., 2014). Together, these findings suggest that the neural representation of language consists of specialized bilateral subnetworks (Cogan et al., 2014) that tune to certain feature characteristics independent of other features within the signal. Crucially, our findings provide evidence for the interaction of auditory, sensorimotor, and somatosensory brain circuitries during speech perception, in conformity with the behavioral link between perception and production faculties in everyday life. The applications of fMRI decoding and generalization methods also hold promise to investigate similarities of acoustic and articulatory speech representations across the perception and production of speech.

References

- Arsenault JS, Buchsbaum BR (2015) Distributed neural representations of phonological features during speech perception. *J Neurosci* 35:634–642. [CrossRef Medline](#)
- Baddeley A, Gathercole S, Papagno C (1998) The phonological loop as a language learning device. *Psychol Rev* 105:158–173. [CrossRef Medline](#)
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312. [CrossRef Medline](#)
- Boersma P, Weenink D (2001) Praat: doing phonetics by computer. *Glott Int* 5:341–347.
- Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014) Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci* 34:4548–4557. [CrossRef Medline](#)
- Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J Neurosci* 34:12662–12677. [CrossRef Medline](#)
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. [CrossRef Medline](#)
- Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G (2011) Conduction aphasia, sensory-motor integration, and phonological short-term memory: an aggregate analysis of lesion and fMRI data. *Brain Lang* 119:119–128. [CrossRef Medline](#)
- Buchweitz A, Shinkareva SV, Mason RA, Mitchell TM, Just MA (2012) Identifying bilingual semantic neural representations across languages. *Brain Lang* 120:282–289. [CrossRef Medline](#)
- Caplan D, Waters GS (1995) On the nature of the phonological output planning processes involved in verbal rehearsal: evidence from aphasia. *Brain Lang* 48:191–220. [CrossRef Medline](#)
- Chen Y, Namburi P, Elliott LT, Heinzl J, Soon CS, Chee MW, Haynes JD (2011) Cortical surface-based searchlight decoding. *Neuroimage* 56:582–592. [CrossRef Medline](#)
- Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014) Sensory-motor transformations for speech occur bilaterally. *Nature* 507:94–98. [CrossRef Medline](#)
- Correia JM, Jansma B, Hausfeld L, Kikkert S, Bonte M (2015) EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Front Psychol* 6:1–10. [CrossRef Medline](#)
- Correia J, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M (2014) Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *J Neurosci* 34:332–338. [CrossRef Medline](#)
- D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. *Curr Biol* 19:381–385. [CrossRef Medline](#)
- De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43:44–58. [CrossRef Medline](#)
- Dick AS, Bernal B, Tremblay P (2014) The language connectome: new pathways, new concepts. *Neuroscientist* 20:453–467. [CrossRef Medline](#)
- Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci U S A* 111:7126–7131. [CrossRef Medline](#)
- Etzel JA, Zacks JM, Braver TS (2013) Searchlight analysis: promise, pitfalls, and potential. *Neuroimage* 78:261–269. [CrossRef Medline](#)
- Evans S, Davis MH (2015) Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cereb Cortex*. Advance online publication. Retrieved Jul. 8, 2015. doi: 10.1093/cercor/bhv136. [CrossRef Medline](#)
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster size threshold. *Magn Reson Med* 33:636–647. [CrossRef Medline](#)
- Formisano E, De Martino F, Bonte M, Goebel R (2008) “Who” is saying “what?” Brain-based decoding of human voice and speech. *Science* 322:970–973. [CrossRef Medline](#)
- Frost MA, Goebel R (2012) Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59:1369–1381. [CrossRef Medline](#)
- Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13:361–377. [CrossRef Medline](#)
- Glasser MF, Rilling JK (2008) DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18:2471–2482. [CrossRef Medline](#)
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp* 27:392–401. [CrossRef Medline](#)
- Guenther FH, Vladusich T (2012) A neural theory of speech acquisition and production. *J Neurolinguistics* 25:408–422. [CrossRef Medline](#)
- Hagler DJ Jr, Saygin AP, Sereno MI (2006) Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* 33:1093–1103. [CrossRef Medline](#)
- Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang Cogn Process* 27:1145–1166. [CrossRef](#)
- Hickok G (2009) Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21:1229–1243. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673–682. [CrossRef Medline](#)
- Hickok G, Okada K, Serences JT (2009) Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101:2725–2732. [CrossRef Medline](#)
- Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422. [CrossRef Medline](#)
- Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum Neurosci* 5:82. [CrossRef Medline](#)
- Jacquemot C, Scott S (2006) What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci* 10:480–486. [CrossRef Medline](#)
- Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31:1715–1720. [CrossRef Medline](#)
- Krieger-Redwood K, Gaskell MG, Lindsay S, Jefferies E (2013) The selective role of premotor cortex in speech perception: a contribution to phoneme judgments but not speech comprehension. *J Cogn Neurosci* 25:2179–2188. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Ladefoged P, Johnson K (2010) A course in phonetics, Ed 6. Boston: Cengage Learning.
- Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J Neurosci* 32:3942–3948. [CrossRef Medline](#)
- Ley A, Vroomen J, Hausfeld L, Valente G, De Weerd P, Formisano E (2012) Learning of new sound categories shapes neural response patterns in human auditory cortex. *J Neurosci* 32:13273–13280. [CrossRef Medline](#)
- Lieberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36. [CrossRef Medline](#)
- Makris N, Preti MG, Asami T, Pelavin P, Campbell B, Papadimitriou GM, Kaiser J, Baselli G, Westin CF, Shenton ME, Kubicki M (2013) Human middle longitudinal fascicle: variations in patterns of anatomical connections. *Brain Struct Funct* 218:951–968. [CrossRef Medline](#)
- Meister I, Wilson S, Deblieck C, Wu A, Locoboni M (2007) The essential role of premotor cortex in speech perception. *Curr Biol* 17:1692–1696. [CrossRef Medline](#)
- Merrill J, Sammler D, Bangert M, Goldhahn D, Lohmann G, Turner R, Friederici AD (2012) Perception of words and pitch patterns in song and speech. *Front Psychol* 3:1–13. [CrossRef Medline](#)
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010. [CrossRef Medline](#)
- Moser D, Baker JM, Sanchez CE, Rorden C, Fridriksson J (2009) Temporal order processing of syllables in the left parietal lobe. *J Neurosci* 29:12568–12573. [CrossRef Medline](#)

- Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci U S A* 108:9998–10003. [CrossRef Medline](#)
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430. [CrossRef Medline](#)
- Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011) A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage* 56:593–600. [CrossRef Medline](#)
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. *Nat Neurosci* 11:367–374. [CrossRef Medline](#)
- Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11:351–360. [CrossRef Medline](#)
- Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci U S A* 103:7865–7870. [CrossRef Medline](#)
- Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740. [CrossRef Medline](#)
- Rietveld ACM, van Heuven VJJP (2001) *Algemene Fonetiek*, Ed 2. Bussum, The Netherlands: Coutinho.
- Schomers MR, Kirilina E, Weigand A, Bajbouj M, Pulvermüller F (2014) Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cereb. Cortex* 25:3894–3902. [CrossRef Medline](#)
- Seltzer B, Pandya DN (1984) Further observations on parieto-temporal connections in the rhesus monkey. *Exp Brain Res* 55:301–312. [Medline](#)
- Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. *Neuroimage* 54:2418–2425. [CrossRef Medline](#)
- Simanova I, Hagoort P, Oostenveld R, van Gerven MAJ (2014) Modality-independent decoding of semantic information from the human brain. *Cereb Cortex* 24:26–434. [CrossRef Medline](#)
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Thieme.
- Wilson S, Saygin A, Sereno M, Lacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702. [CrossRef Medline](#)
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849. [CrossRef Medline](#)