

Title: Categorical speech processing in Broca's area: An fMRI study using  
multivariate pattern-based analysis

Running title: Categorical speech processing in Broca's area

Authors: Yune-Sang Lee (이윤상)<sup>1</sup>, Peter Turkeltaub<sup>2</sup>, Richard Granger<sup>1</sup>, and  
Rajeev D. S. Raizada<sup>3</sup>

Institute: <sup>1</sup> Dept. of Psychological and Brain Sciences, Dartmouth College,  
Hanover, NH

<sup>2</sup> Neurology Department, Georgetown University, Washington, DC

<sup>3</sup> Neukom Institute, Dartmouth College, Hanover, NH

Corresponding author: Yune-Sang Lee, Ph.D

Center for Cognitive Neuroscience

3710 Goddard laboratory, Hamilton Walk

University of Pennsylvania

Philadelphia, PA, 19104

E-mail: [yslee@mail.med.upenn.edu](mailto:yslee@mail.med.upenn.edu)

Number of pages: 18 (including a title page)

Number of figures: 3 / Number of tables: 1

Number of words for abstract: 228

Number of words for introduction: 507

Number of words for discussion: 1548

Total number of words: 5520

There is no conflict of interest

Acknowledgment: The authors thank Yale Cohen, Tilbe Goksun, David Kraemer, Kate Onsgard for their helpful comments on the earlier version of the manuscript. Special thanks go to the Penn postdoc editors association for proof-reading. This work was supported in part by grants from the Office of Naval Research.

Rajeev Raizada is currently affiliated with Dept. of Human Development, Cornell University, Ithaca, NY.

## **Abstract**

Although much effort has been directed towards understanding the neural basis of speech processing, the neural processes involved in the categorical perception of speech have been relatively less studied, and many questions remain open. In this functional magnetic resonance imaging (fMRI) study, we probed the cortical regions mediating categorical speech perception using an advanced brain-mapping technique, whole-brain multivariate pattern-based analysis (MVPA). Normal healthy human subjects (native English speakers) were scanned while they listened to each of 10 consonant-vowel syllables along the /ba/-/da/ continuum. Outside of the scanner, individuals' own category boundaries were measured in order to divide the fMRI data into /ba/ and /da/ conditions per subject. The whole-brain MVPA revealed that Broca's area and the left pre-SMA (supplementary motor area) evoked distinct neural activity patterns between the two perceptual categories (/ba/ vs. /da/). Broca's area was also found when the same analysis was applied to another data set (Raizada and Poldrack, 2007), which previously yielded the supramarginal gyrus (SMG) using a univariate adaptation-fMRI paradigm. The consistent MVPA findings from two independent data sets strongly indicate that Broca's area participates in categorical speech perception, with a possible role of translating speech signals into articulatory codes. The difference in results between univariate and multivariate pattern-based analyses of the same data suggest that processes in different cortical areas along the dorsal speech perception stream are distributed on different spatial scales.

## 1 Introduction

One of the core aspects of speech perception is mapping complex time-varying acoustic signals into discrete speech units. The “categorical speech perception” phenomenon was experimentally demonstrated in the 1950s, whereby synthetic syllables along the continuum between prototypes (e.g., /ba/ vs. /da/) were perceived categorically despite their linear acoustic variations (Liberman et al., 1957). According to one of the most influential speech theories (Liberman and Mattingly, 1985), this perceptual categorization of incoming auditory speech occurs because articulatory gestures serve as the brain’s representations of speech sounds, and speech is perceived by mapping continuous auditory signals onto discrete articulatory gestures. More recent theories based on converging evidence from lesion, neuroimaging and brain stimulation studies (for a review, see Devlin and Aydelott, 2009) discount the role of articulatory gestures as perceptual representations, postulating instead that speech percepts and motor codes interact via feedforward and feedback connections (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

This auditory-motor integration is thought to be achieved along a dorsal stream speech network, running from primary auditory cortex via posterior superior temporal gyrus (STG) and the inferior parietal lobule to the posterior frontal lobe (Hickok and Poeppel, 2007). The left posterior STG has been implicated in categorical phoneme perception by functional imaging (Husain et al., 2006; Desai et al., 2008; Hutchison et al., 2008) and intracranial human electrophysiology studies (Chang et al., 2010). The SMG has been also implicated in categorical speech perception. Raizada and Poldrack (2007) found that the SMG evoked amplified neural activity in response to between-category phoneme pairs along the syllabic continuum of /ba/ and /da/. Furthermore, a recent meta-analysis revealed that the SMG and angular gyrus were consistently activated in fMRI studies of categorical speech perception (Turkeltaub and Coslett, 2010). Thus, posterior nodes of the dorsal speech pathway involved in spectrotemporal analysis of auditory signals, phonological processing, and sensorimotor interface have been clearly implicated in categorical perception of speech. In contrast, findings in frontal articulatory coding areas (posterior

inferior frontal gyrus, premotor cortex, anterior insula) have occasionally been reported in studies of categorical speech perception (Myers et al., 2009), but are far less consistent.

We hypothesized that frontal articulation areas are involved in categorical speech perception, but that they may be invisible to subtraction-based fMRI analysis if complex articulatory gestures are represented not by different levels of activity within single voxels, but by differential neural activity patterns within a region of cortex. As such, the present fMRI study employed a relatively new mapping strategy, the searchlight analysis (Kriegeskorte et al., 2006), to explore the neuroanatomical basis of categorical speech perception. This technique, based upon multivariate pattern-analysis (MVPA) (Haxby et al., 2001; Pereira et al., 2009), probes local information of neural pattern differences across different conditions by moving a searchlight unit on a voxel-by-voxel basis. Some fMRI studies have used MVPA to examine speech perception (Formisano et al., 2008; Okada et al., 2010; Raizada et al., 2010; Kilian-Hutten et al., 2011). Most, however, examined the neural activity patterns exclusively within superior temporal regions. The searchlight procedure allowed us to examine the entire brain.

## **2 Materials and Methods**

### *2.1 Subjects*

Fifteen right-handed native English speakers without reported hearing difficulties or neurological disorders (10 males/5 females, ages 19-34 years) were recruited from the Dartmouth College community. Two were excluded (one fell asleep during scanning, and one did not have a clear categorical boundary in the behavioral experiment), leaving thirteen subjects total. Consent was obtained from all subjects as approved by the Committee for the Protection of Human Subjects at Dartmouth College.

### *2.2 Stimuli*

Ten synthesized phonemes (duration of each: 300 ms) along the /ba/-/da/ continuum were created by varying the second and third formant using a

SenSyn Klatt Synthesizer (Sensimetrics, Inc.), as described in Raizada and Poldrack (2007). Each phoneme along the continuum was referred in number with “1” being the stimulus of clear /ba/ and “10” being /da/ (Figure 1).

## *2.3 Experimental Procedures*

### 2.3.1 fMRI experiment

During scanning, subjects listened to the stimuli while performing a non-phonetic task designed to maintain alertness, in which they were required to indicate a quieter stimulus that was presented in the catch trial block by pressing a button with their right hand.

### 2.3.2 Behavioral experiment

After the fMRI experiment, subjects were behaviorally tested to measure a categorical boundary on the 10 syllabic tokens that were previously presented in the scanner. Each syllabic sound was presented eight times, and subjects were required to indicate if they perceived /ba/ or /da/ by button-press. The boundary was defined as the 50% crossover point between perception of /ba/ and /da/ on each subject’s psychometric curve and was used to define binary classes (/ba/ vs. /da/) to label neural data per subject.

## *2.4 fMRI scanning*

A Philips Intera Achieva 3T scanner was used with a standard EPI BOLD pulse sequence and a clustered volume acquisition. The parameters are as follows: TR=3s; TA=1.88s; silent gap = 1.12s; 560ms interval between stimuli and scanner-noise onset/offset; 32 slices; 3x3mm within plane resolution, 4mm thick slices with a 0.5mm skip; interleaved slice-ordering. Each stimulus was presented in the middle of the silent gap prior to the next onset of EPI acquisition. In the scanner, sounds were played via high-fidelity MR compatible headphones (MR Confon, Germany). The fMRI scans were subdivided into 5 runs with 185 volume acquisitions per run. A block design was used, with one of the 10 phonemes repeatedly presented 5 times in each block. Between the blocks were rest periods lasting 15s (5 TRs). The order of the blocks was pseudo-random and counter-balanced across subjects. There were 18 stimulus blocks per run,

making 90 blocks total across the 5 runs. Of these, 10 contained quieter catch-trials which were excluded from subsequent analysis. The remaining 80 blocks consisted of 8 blocks for each of the 10 stimuli along the /ba/-/da/ continuum.

### *2.5.1 fMRI data (searchlight procedure) analysis methods*

fMRI data were preprocessed using SPM5 and MATLAB 2009b. All images were realigned to the first EPI, and spatially normalized into Montreal Neurological Institute (MNI) standard stereotactic space (ICBM152 EPI template) with preserved original voxel size. After preprocessing, fMRI time courses of all voxels were extracted from unsmoothed images. Subsequently, these raw signals were high-pass filtered with a 300s cut-off to remove slow drifts caused by the scanner, and standardized across entire runs using the z-score to normalize intensity differences among runs. When the data were subsequently entered into a classifier in pattern-based analysis, the TRs corresponding to each syllabic stimulus were labeled as /ba/ or /da/ by using the stimulus blocks' presentation times, as follows. For each stimulus-type, the time courses of its stimulus-presentation blocks were convolved with a canonical HRF, thereby producing a regressor for that class of stimuli. The mean value across all time points of the regressor was then calculated. A time-point was assigned as belonging to the given stimulus-class if the value of the regressor at that time-point was greater than this overall mean value.

For the primary searchlight analysis, local searchlight spheres consisting of a center voxel and its neighborhood within three-voxel radius were constructed throughout the entire brain (approximately 30,000 units). Then, in each searchlight sphere, a binary classification was performed using the Gaussian Naïve Bayes (GNB) classifier (Pereira et al., 2009). The classifier was initially trained by a subset of data sets (training set) to build a model that appropriately set the boundary between the neural activities that were associated with a perceptual category of /ba/ and /da/. Then, this model was applied to the remaining data sets (testing set) for validation, wherein the accuracy of classification test was gained by calculating how many times the classifier correctly predicted time-points corresponding either to /ba/ or to /da/ in the

unseen data set (chance level=50%). To avoid a bias caused by a particular training/ testing set, this procedure was repeated five times, such that signals from four scanning runs were served as a training set and one remaining run served as a testing set (i.e., five-fold cross-validation). The classification accuracy for each searchlight sphere was averaged across the testing-set results of the five training/ testing combinations and stored in each voxel of an output image for each subject (for more details, see Krigeskorte et al. 2006).

After output images were acquired by searchlight analysis, baseline-corrections in voxel values were made via two steps. First, the chance-level accuracy, i.e., 0.5 was subtracted from an accuracy stored in each voxel of the output images. Subsequently, the mean of these accuracy values in the output image was subtracted from each voxel's value, such that the resulting output image for each subject had a mean of zero. This baseline-correction procedure has two advantages. First, this ensures to guard against false discovery (i.e., more conservative than the standard procedure with no baseline correction). Secondly, this helps normalizing different accuracy ranges across subjects (i.e., reducing the variance in the signal at a group level, which in turn increases the t-statistical power at the group level). The adjusted accuracy map was submitted to the random effects analysis and threshold of  $p$  (*cluster-wise corrected*)  $< 0.05$  in combination with  $p$  (*voxel-wise uncorrected*)  $< 0.001$  was applied throughout the searchlight analyses. For the random effects analyses on both current and old data sets, SPM8 was used instead in order to take advantage of its newly added feature, cluster-level correction based upon FDR (False Discovery Rate) in the updated version of SPM.

### 2.5.2 A new analysis, using a searchlight approach, of the Raizada & Poldrack 2007 data set

Procedures were identical to 2.5.1 except that the smoothed data of 12 subjects were used because unsmoothed data were not available (For details of data, see Raizada & Poldrack 2007). In that previous study, stimuli were presented in pairs of two types using event-related design: identical pairs, where one token on the 10-step /ba/-/da/ continuum was presented twice in succession (e.g. 4-then-4), and 3-step pairs, in which the two stimuli were three

steps apart along the continuum (e.g. 4-then 7). In the present analysis, only the identical pairs were analyzed, as these are directly comparable to the uniform blocks of stimuli that were used in the newly acquired data.

### *2.5.3 Conventional univariate GLM analysis*

For comparison purposes, we also carried out standard univariate GLM analyses of the data, investigating the following four contrasts: /ba/ minus /da/, /da/ minus /ba/, and as simple controls, all-speech-minus-baseline and baseline-minus-all-speech. These GLM analyses were applied to the smoothed images of the current data set for each subject using SPM5. At the first level analysis, the fMRI time-courses corresponding to each block of stimulus were convolved with canonical hemodynamic response function (HRF) using the box-car design. Additionally, the six motion parameters, as well as the onset of button responses were modeled, which were later regressed out as effects of non-interest. Contrast images from each individual subject's first-level analysis were submitted to the 2<sup>nd</sup> level random effects for group analysis.

## **3 Results**

### **3.1 Behavioral results**

As expected, nearly all subjects demonstrated a sharp category boundary near the middle of the syllabic continuum between /ba/ and /da/ (Figure 1). However, one subject did not perceive the syllabic continua in a categorical manner (i.e., linear trend) and was later discarded from further analysis. This subject was a native English speaker, but spent several years during childhood in Hong Kong, which may have affected the behavioral results. Indicated by error bars in Figure 1, the category boundary varied slightly among the subjects.

### **3.2 fMRI**

#### *3.2.1 MVPA*

The whole-brain searchlight analysis equipped with three-voxel radius



spheres revealed two left frontal regions, namely, the pre-SMA and the pars opercularis within Broca's area (Figure 2, Figure 3a, Table1). No cortical regions within the temporal or parietal lobes were significant. In order to further validate these results, we re-analyzed previously published adaptation fMRI data (Raizada and Poldrack, 2007) using the same MVPA procedures. Despite several differences between these two data sets (design, number of available trials, degree of smoothness, participating subjects), the left pars opercularis again was significant at a matched threshold (Figure 3c, Table 1); no other significant areas were identified. Note that the previous data set implicated the left SMG in categorical phoneme perception using the adaptation fMRI paradigm (Figure 3d, Table1). However, this area did not emerge when the MVPA procedure was used on either data set ( $t_{12} = 1.3$ ,  $t_{11} = 0$ ,  $p = n.s.$ ; for new and old data set respectively)

To confirm that the activity identified in the MVPA analyses related to individual subjects' percepts rather than acoustic differences between items at each end of the continuum, we performed another searchlight analysis with categories binarized based on the absolute mid-point along the continuum, disregarding each subject's category boundary. As a result, no significant areas were found even at a relaxed threshold ( $p$  (voxel-wise uncorrected)  $< 0.01$ ).

The difference between the searchlight analysis results and the previous univariate results could be interpreted as a difference in the spatial scale of processing in different areas of the network (intra-voxel in the SMG vs. inter-voxel in the pars opercularis). This idea raised the possibility that processing in other areas of the speech perception network might be distributed over regions of cortex too large to be detected using a 3-voxel searchlight. To explore such possibility, we performed additional searchlights analyses with a systematic manipulation of the searchlight size parameter. At a radius size of eight-voxels, a significant cluster emerged within the left STG in the new data set at a matched threshold (figure 3d, Table1). Interestingly, the pars opercularis and pre-SMA were no longer significant in that larger searchlight analysis. Nevertheless, old data set did not yield significant temporal clusters in any of those additional analyses.

### 3.2.2 Conventional univariate analysis

The comparison of all sounds vs. resting blocks yielded significant clusters within the expected cortical regions along the speech network including bilateral temporal lobes, and left parietal and frontal region (figure not shown). By contrast, a complementary univariate comparison of items perceived by each subject as /ba/ vs. /da/ revealed no significant areas even at a relaxed threshold ( $p$  (voxel-wise uncorrected)  $< 0.01$ ), confirming that MVPA clusters did not represent a simple difference in voxel-wise activation intensity.

#### 4 Discussion

Although the neural basis of speech perception has been extensively studied in the auditory neuroimaging field, relatively little evidence has been garnered as to how speech sounds are sorted into discrete categories (Türkeltaub and Coslett, 2010). Previous evidence, based largely on standard univariate fMRI analysis, has mainly suggested roles for posterior temporal and inferior parietal regions in categorical speech perception (Celsis et al., 1999; Zevin and McCandliss, 2005; Husain et al., 2006; Raizada and Poldrack, 2007; Desai et al., 2008; Hutchison et al., 2008; Chang et al., 2010; Zevin et al., 2010; Kilian-Hutten et al., 2011). Nevertheless, theories implicating sensorimotor mapping in speech sound categorization (Devlin and Aydelott, 2009) suggest frontal regions should be involved as well. In line with such notion, the present MVPA study revealed that two frontal region, namely, Broca's area and left pre-SMA evoked neural activities related to perceptual category of syllabic stimuli. The same Broca's area result was observed when the identical MVPA procedure was applied to another data set that had previously implicated the SMG as an important hub of categorical speech processing (Raizada and Poldrack, 2007). The frontal network has been typically viewed as a region involved in speech production and so, our findings naturally invite consideration of auditory-motor integration in speech perception.

##### *The possible roles of Broca's area in speech perception*

Converging evidence from multiple sources has demonstrated that there

exists an anterior-posterior distinction within Broca's area, where anterior areas play a role in semantic processes and posterior are involved in phonetic and phonological processes (Poldrack et al., 1999; Gold et al., 2005; Gough et al., 2005). Lesion data as well as functional imaging results have suggested a role for pars opercularis (the posterior-most region of Broca's area) in sublexical phonetic processing during speech production (Carreiras et al., 2006; Riecker et al., 2008; Papoutsis et al., 2009). The pars opercularis was the significant region of Broca's area in the present study as well. Thus it is possible that the same phonetic processes or representations that operate in Broca's area during speech production or decision tasks also participate in sublexical speech perception.

Another possible role for posterior Broca's area in speech perception relates to its role in motor aspects of speech, specifically articulatory planning. Recently, evidence from neuroimaging (Wilson et al., 2004) and TMS studies (Watkins et al., 2003) has suggested that motor systems play a role in speech perception (Pulvermüller et al., 2006; Devlin and Aydelott, 2009). These findings have been also viewed as evidence supporting putative mirror neuron systems in the human brain (D'Ausilio et al., 2009). Relatedly, Skipper et al., (2007) presented evidence that visual contexts (e.g., viewing mouth movement) aid speech perception by recruiting motor systems. Consistent with that work, the present study found that a frontal motor-speech network consisting of Broca's area and the left pre-SMA evoked distinct neural patterns corresponding to the perceived syllables. One plausible scenario is that Broca's area generated articulatory representations, in order to guide categorization of the phonetic continua into discrete syllabic units via feedback mechanisms, as proposed in various current speech models (Callan et al., 2004; Davis and Johnsrude, 2007; Rauschecker and Scott, 2009). This articulatory guidance was presumably subconscious in these experiments, as the fMRI task was orthogonal to syllabic identity and subjects were not aware of the behavioral categorization task until afterwards.

We also identified pre-SMA activity related to syllable identity. The pre-SMA, which rarely if ever activates in univariate imaging studies of speech perception, has been viewed as a high-level action control region mediating sensory-motor transition (Nachev et al., 2008), and has recently received

attention in speech production (Alario et al., 2006). Lesions in pre-SMA cause deficits in production of planned speech (Ziegler et al., 1997), suggesting that it plays a role in the preparation of sequential movements, especially in conditions involving a short-term buffering of motor speech acts. Thus, the pre-SMA finding here further supports involvement of articulatory motor processes in speech perception (Pulvermuller et al., 2006). We surmise that articulatory codes and sequences accessed in Broca's area and nearby premotor cortex may be transmitted to the pre-SMA for buffering during speech production or modeling to guide speech perception. In particular, the frontal motor system may play a role when identifying ambiguous or degraded speech signals. For example, a neuroimaging study revealed that, during /r/ vs. /l/ consonant identification task, Broca's area was more activated in Japanese speakers whose mother tongue did not differentiate the two consonant sounds than in native English speakers (Callan et al., 2004). Motor systems may also contribute to perception in adverse listening circumstances, as evidenced by the specific perceptual difficulty patients with Broca's aphasia experience when speech signals are degraded (Utman et al., 2001; Moineau et al., 2005). Together, the current findings lend further support for the notion that auditory and motor systems interact during speech perception.

*Different spatial scales of processing along the dorsal speech perception stream*

Intriguingly, the MVPA and univariate adaptation analysis of the same data set (Raizada and Poldrack, 2007) yielded completely different results, the former identifying Broca's area, and the latter SMG. For this new analysis, only the within-category pairs were chosen for binary classification whereas, in the previous study, the neural activities of between-category pairs were contrasted with those of within-category pairs. Thus, it is important to underscore that the analyses were not performed for identical comparisons, although both were designed to isolate the same perceptual processes. Recently Zevin and colleagues (2010) demonstrated that the SMG is involved in domain-general change detection, and is thus commonly active during adaptation paradigms regardless of speech content. Hence, the discrepancy between the univariate and multivariate results may reflect different roles for SMG (change detection or

discrimination) and Broca's area (phoneme identification) in speech perception.

However, in the visual domain, simultaneous application of a univariate adaptation paradigm (Grill-Spector and Malach, 2001) and a multivariate analysis on the identical data (Drucker and Aguirre, 2009) has also revealed differences in localization of activity, as in our study. This difference may have occurred because local networks in different areas of the brain operate at different spatial scales reflecting different underlying computational properties. Different types of fMRI analysis reveal different spatial scales of neural processing. Adaptation-fMRI distinguishes between differently-tuned populations of neurons that are intermingled within the same voxel, and so it highlights processing at a sub-voxel spatial scale. In contrast, MVPA reveals areas where information is represented by distributed patterns of activity, thereby it highlights processing at a multi-voxel spatial scale. Thus, the findings here may indicate that Broca's area and the SMG utilize different spatial scales (inter vs. intra-voxel, respectively) to process speech inputs. Recently, MVPA studies have shown distributed neural patterns for speech perception within large expanses of the temporal lobes (Formisano et al., 2008; Kilian-Hutten et al., 2011). One possible reason for the discrepancy between our main findings and these studies is that auditory representations of phonemes are coded by patterns of activity distributed across large expanses of temporal cortex, while frontal representations are coded in patterns across smaller expanses. Thus, the relatively small three-voxel searchlight might be spatially tuned to detect frontal representations, but too small to detect temporal representations. To test this idea, we performed additional searchlight analyses with larger sizes of local spheres. At a radius of eight voxels, the pars opercularis cluster was no longer significant, but instead the searchlight revealed a significant cluster in the left posterior STG in the new data set. This finding supports the notion that frontal and temporal speech representations are distributed at different spatial scales. Because of the large size of the searchlights in this analysis (approximately 21 mm radius), searchlights contributing to this STG cluster included voxels in the mid-to-posterior STG and MTG, and portions of the SMG also. This could be taken as a caveat to our claim above that SMG speech processing occurs at a small spatial scale. However, much of the STG cluster was too far from the SMG

to have included it in the searchlights. The failure to replicate this finding in the old data set should also be taken as a caveat, but was likely because unsmoothed images were no longer available and only subsets of the trials in the old dataset were used. However, future studies should also consider alternative explanations for differential spatial scale along the fronto-temporal speech network, such as differing hemodynamic response profiles across different regions (Handwerker et al., 2004). Because MVPA are sensitive to coherent patterns of information-carrying activity rather than to univariate intensity differences, it is unlikely that differing hemodynamic response profiles across different regions could account for the current findings. Nonetheless, it would be interesting for future studies to investigate these questions further.

In summary, our results clearly demonstrate the presence of phoneme category-specific information in pars opercularis within Broca's area. Further research is needed to clarify the specific role of this area in speech perception. Our findings also raise interesting questions regarding differences in the spatial scale of cortical representations or processes along speech perception pathways. This principle likely applies to other perceptual domains as well. The possibility that different regions of the cortex encode information at different scales of spatial distribution, or even that a single region may encode different information at different spatial scales should be considered and examined further in future studies.

## References

- Alario FX, Chainay H, Lehericy S, Cohen L (2006) The role of the supplementary motor area (SMA) in word production. *Brain Res* 1076:129-143.
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22:1182-1194.
- Carreiras M, Mechelli A, Price CJ (2006) Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Hum Brain Mapp*

- 27:963-972.
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F (1999) Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9:135-144.
- Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428-1432.
- D'Ausilio A, Pulvermuller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. *Curr Biol* 19:381-385.
- Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear Res* 229:132-147.
- Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci* 20:1174-1188.
- Devlin JT, Aydelott J (2009) Speech perception: motoric contributions versus the motor theory. *Curr Biol* 19:R198-200.
- Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb Cortex* 19:2269-2280.
- 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.
- Gold BT, Balota DA, Kirchoff BA, Buckner RL (2005) Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from fMRI adaptation. *Cereb Cortex* 15:1438-1450.
- Gough PM, Nobre AC, Devlin JT (2005) Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J Neurosci* 25:8010-8016.
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107:293-321.
- Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21:1639-1651.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425-2430.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393-402.
- Husain FT, Fromm SJ, Pursley RH, Hosey LA, Braun AR, Horwitz B (2006) Neural bases of categorization of simple speech and nonspeech sounds. *Hum Brain Mapp* 27:636-651.
- Hutchison ER, Blumstein SE, Myers EB (2008) An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage* 40:342-352.
- Kilian-Hutten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31:1715-1720.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *P Natl Acad Sci USA* 103:3863-3868.
- Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised.

- Cognition 21:1-36.
- Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The Discrimination of Speech Sounds within and across Phoneme Boundaries. *J Exp Psychol* 54:358-368.
- Moineau S, Dronkers NF, Bates E (2005) Exploring the processing continuum of single-word comprehension in aphasia. *J Speech Lang Hear Res* 48:884-896.
- Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol Sci* 20:895-903.
- Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. *Nat Rev Neurosci* 9:856-869.
- Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G (2010) Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex* 20:2486-2495.
- Papoutsis M, de Zwart JA, Jansma JM, Pickering MJ, Bednar JA, Horwitz B (2009) From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cereb Cortex* 19:2156-2165.
- Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage* 45:S199-S209.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15-35.
- Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. *P Natl Acad Sci USA* 103:7865-7870.
- Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726-740.
- Raizada RD, Tsao FM, Liu HM, Kuhl PK (2010) Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. *Cereb Cortex* 20:1-12.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718-724.
- Riecker A, Brendel B, Ziegler W, Erb M, Ackermann H (2008) The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain Lang* 107:102-113.
- Skipper JL, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387-2399.
- Turkeltaub PE, Coslett HB (2010) Localization of sublexical speech perception components. *Brain Lang* 114:1-15.
- Utman JA, Blumstein SE, Sullivan K (2001) Mapping from sound to meaning: reduced lexical activation in Broca's aphasics. *Brain Lang* 79:444-472.
- Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41:989-994.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates



motor areas involved in speech production. *Nat Neurosci* 7:701-702.

Zevin JD, McCandliss BD (2005) Dishabituation of the BOLD response to speech sounds. *Behav Brain Funct* 1:4.

Zevin JD, Yang J, Skipper JI, McCandliss BD (2010) Domain general change detection accounts for "dishabituation" effects in temporal-parietal regions in functional magnetic resonance imaging studies of speech perception. *J Neurosci* 30:1110-1117.

Ziegler W, Kilian B, Deger K (1997) The role of the left mesial frontal cortex in fluent speech: evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia* 35:1197-1208.

## Figure Legends

**Figure 1.** The average identification curve from the 13 participants. The x-axis shows the 10-step syllabic continuum in numbers between 1 (/ba/ end point) and 10 (/da/ end point). The acoustic distance between successive steps is identical. The y-axis shows the percentage of /da/ response frequency at each point and the error bars indicate standard errors.

**Figure 2.** Broca's area (MNI  $x,y,z = -54, 12, 22$ ) and pre-SMA (MNI  $x,y,z = -3, 18, 50$ ) clusters were overlaid onto the surface inflated rendering brain using SPM surfrend toolbox (<http://spmsurfrend.sourceforge.net/>).

**Figure 3. Top panel depicts results from new data set a.** Broca's area (MNI  $x,y,z = -54, 12, 22$ ) and **b.** left STG/S identified by the searchlight analyses equipped with 3-voxel and 8-voxel radius of local spheres respectively. **Bottom panel depicts results from another data set. c.** Broca's area (MNI  $x,y,z = -45, 9, 27$ ) identified by the searchlight analysis with 3-voxel radius of local sphere. **d.** The left SMG (MNI  $x,y,z = -60, -24, 18$ ) previously identified by adaptation analysis (Raizada & Poldrack, 2007). The color bars indicate t-values of the group random effects analysis.

Table 1

## Cortical loci involved in categorical speech perception

Region name	MNI Coordinates			z-value	t-value	cluster	Data set		
	HEM	BA	x y z						
<i>MVPA (3-voxel radius searchlight)</i>									
PreSMA	L	6	-3	18	50	4.32	6.94	12	Current fMRI data
Pars Opercularis	L	44	-54	12	22	4.09	6.25	24	Current fMRI data
Pars Opercularis	L	44	-45	9	27	3.68	5.35	25	Raizada & Poldrack (2007)
<i>MVPA (8-voxel radius searchlight)</i>									
Superior Temporal Gyrus/Sulcus	L	22	-51	-36	23	4.21	6.60	42	Current fMRI data
<i>Adaptation</i>									
Supramarginal Gyrus	L	40	-60	-24	18	4.78	9.16	30	Raizada & Poldrack (2007)

The statistical maps were significant at voxel-wise uncorrected  $p < 0.001$  & cluster-wise corrected  $p < 0.05$ .

HEM = hemisphere; BA = approximate Brodmann area

**Figure 1**

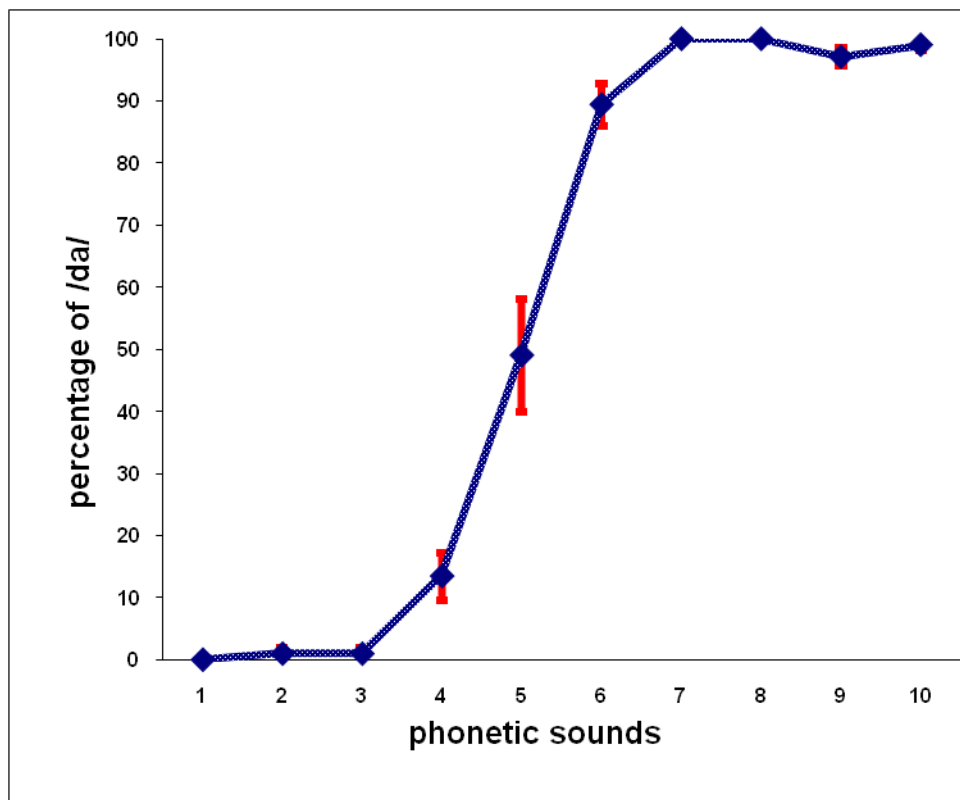


Figure 2

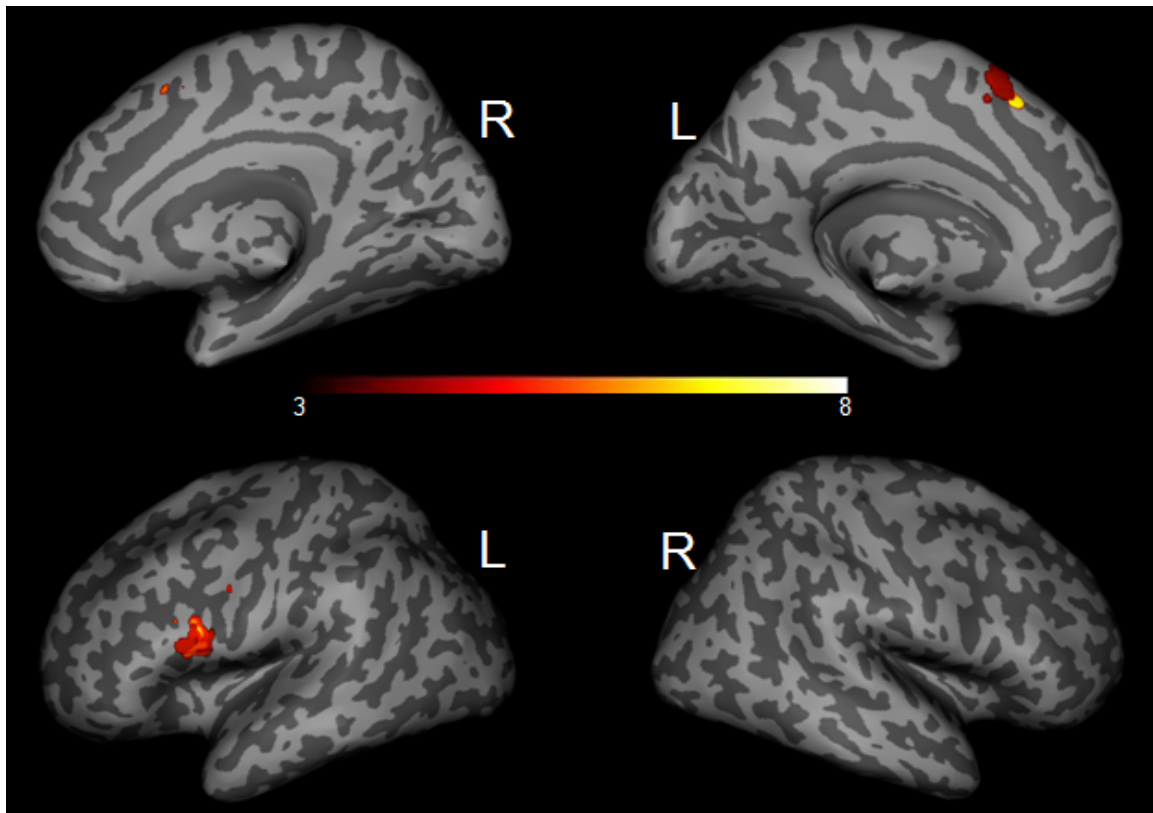


Figure 3

