

Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences

In order for stimuli to be perceptually discriminable, their representations in the brain must be distinct. Investigating the task of discriminating the syllables /ra/ and /la/, we hypothesized that the more distinct a person's neural representations of those sounds were, the better their behavioral ability to discriminate them would be. Standard neuroimaging approaches are ill-suited to testing this hypothesis as they have problems differentiating between neural representations spatially intermingled within the same brain area. We therefore performed multi-voxel pattern-based analysis of the functional magnetic resonance imaging (fMRI) activity elicited by these syllables, in native speakers of English and Japanese. In right primary auditory cortex, the statistical separability of these fMRI patterns predicted subjects' behavioral ability to tell the sounds apart, not only across groups but also across individuals. This opens up a new approach for identifying neural representations and for quantifying their task suitability.

Keywords: auditory cortex, classifier, nonnative language, pattern-based fMRI, SVM

Introduction

During early childhood, people's neural representations of speech sounds become finely tuned to the phonetic distinctions of their own native language (Kuhl 2004). Thus, English speakers' representations are well suited for hearing distinctions that are crucial for the English language, and Japanese speakers' representations are correspondingly well suited for their own mother tongue. However, neural representations that work well for the Japanese language meet with less success when confronted with the task of perceiving English language phonetic contrasts, as the well-known difficulty hearing the difference between /r/ and /l/ attests (Underbakke et al. 1988; Logan et al. 1991; Guion et al. 2000; McCandliss et al. 2002; Iverson et al. 2003). English speakers' representations, in turn, are inadequate for other tasks, such as for hearing the difference between the retroflex and dental /r/ sounds in Hindi (Pruitt et al. 2006).

The present study asks the following question: can we specify how the neural representations of the syllables /ra/ and /la/ differ between English and Japanese speakers, and moreover can we predict individual differences in people's perceptual ability to discriminate those speech sounds, purely on the basis of the neural activation patterns that the sounds evoke in their brains? Specifically, it seeks to test the hypothesis that the more distinct a person's neural representations of /ra/ and /la/ are, the better their behavioral ability to discriminate the 2 syllables will be.

Rajeev D. S. Raizada¹, Feng-Ming Tsao², Huei-Mei Liu³ and Patricia K. Kuhl⁴

¹Neukom Institute for Computational Science, Dartmouth College, Hanover, NH 03755, USA, ²Department of Psychology, National Taiwan University, Taipei 10617, Taiwan, ³Department of Special Education, National Taiwan Normal University, Taipei 10644, Taiwan and ⁴Institute for Learning and Brain Sciences, University of Washington, Seattle, WA 98195, USA

Relations between Neural Activity, Behavioral Performance, and Perceptual Discriminability

Many studies have demonstrated correlations between neural activity and levels of behavioral performance. In most cases, increasing levels of neural activity in a given brain area correlate with better behavioral performance in the task that the region subserves. This type of relation between neural activity and behavioral performance is schematically illustrated in Figure 1A.

In the present study, we take a different approach. Greater intensity of neural activity is not necessarily better: success or failure in a task may be determined not by the intensity of neural activation evoked but instead by the properties of the neural representations that are involved.

This is especially true for discrimination tasks: in order to make a fine-grained behavioral discrimination, the brain needs correspondingly fine-grained neural representations. Thus, the fact that native English speakers have no difficulty hearing the difference between /r/ and /l/ implies that the neural activation patterns that those sounds evoke in their brains must be distinguishable. The hypothesis that we test in the present study predicts that in Japanese speakers, the neural activation patterns evoked by those sounds will not be as distinguishable from each other as they are in English speakers.

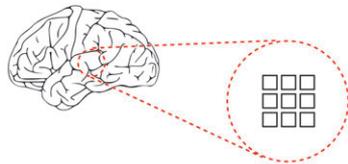
Studies in which more intense neural activity does correspond to improved behavioral performance, unlike in the experiment presented here, are of 2 sorts. The improved performance can be within individuals, with correct trials producing greater activation than incorrect trials, for example, studies of memory encoding in the hippocampus (Brewer et al. 1998; Wagner et al. 1998). Alternatively, the differences in performance can be across individuals, with greater activity being found in people who perform well on a given task compared with people who do not. This type of result has been found across multiple domains: For example, greater neural activity can predict better language ability (Demb et al. 1997; Crinion and Price 2005), mathematical ability (Grabner et al. 2007), or expertise in visual object recognition (Gauthier et al. 2000). In such cases, the more active the brain area is, the more work it gets done.

However, this is not always the case. One example of where more intense activation is not necessarily better is in performing visual acuity tasks. Visual acuity is better at the fovea than in the periphery because the fovea is represented by neurons with smaller receptive fields and not because those neurons fire any more vigorously. Because each cortical neuron represents a smaller patch of visual space at the fovea, the total amount of cortical tissue representing that high-acuity part of the visual field is correspondingly larger. This relation between the size of a cortical representation and behavioral ability in acuity tasks was demonstrated using functional

magnetic resonance imaging (fMRI) by Duncan and Boynton (2003, 2007) and is sketched in Figure 1B.

Because there is a well-defined spatial map in primary visual cortex, it is possible directly to measure how fine grained its neural representations are and to relate this measurement to behavior. However, for speech sounds the question of what a fine-grained neural representation might look like is less clear.

How does the neural activity in a brain area relate to a person's level of behavioral performance?



A Increased neural activity gives better behavioral performance (e.g. Demb et al., 1997; Crinion & Price, 2005)

Amount of activation:	Low	Medium	High
Behavioral performance:	Poor	Moderate	Good

B Larger cortical representation gives better behavioral performance (e.g. Duncan & Boynton, 2003, 2007)

Size of cortical representation:	Small	Medium	Large
Behavioral performance:	Poor	Moderate	Good

C New hypothesis, for discrimination tasks:
If the fMRI activation patterns are more separable, then the evoking stimuli are more perceptually discriminable.

Activation pattern for /ra/			
Activation pattern for /la/			
Distinctness of activation patterns:	Low	Medium	High
Behavioral performance:	Poor	Moderate	Good

Total amount of local activation is the same, across all these patterns

Figure 1. Different varieties of relations between neural activity and behavioral performance. (A) Many previous studies have found circumstances in which greater neural activity in a given brain area is associated with better behavioral performance. (B) In acuity tasks, behavioral performance is associated not with increased activation intensity but instead with the size of the cortical representations of the stimuli. Larger cortical area corresponds to small receptive fields, so that the neural responses are more fine grained. This relation was demonstrated using fMRI by Duncan and Boynton, first in visual cortex (Duncan and Boynton 2003) and subsequently in somatosensory cortex (Duncan and Boynton 2007). (C) In the present study, we test the hypothesis that subjects' behavioral ability to discriminate /ra/ from /la/ depends upon how distinct from each other the underlying neural representations are. The degree of distinctness is not determined at the single-voxel level but instead is measured in terms of how separable the multi-voxel spatial patterns of evoked fMRI activity are. Note that all the activation patterns shown in panel (C) contain equal total amounts of local activation. Thus, this hypothesized relation between neural activity and levels of behavioral performance would be completely different from the "more activation is better" scenario sketched in panel (A).

One way in which 2 neural activation patterns can be distinguished from each other is if they are spatially separated. However, there is no known "phonotopic" map in the brain, for example, with /r/-evoked activation being more anterior and /l/ more posterior. Another way for the activations to be distinguishable would be via intensity, for example, if /r/ were to elicit activation twice as strong as /l/. Again, however, neuroimaging studies suggest that this is not the case (Callan et al. 2003, 2004; Tricomi et al. 2006). Such a coding scheme would also run into difficulties as 2 /l/ stimuli heard together would produce neural activation easily confusable with the activation from a single /r/.

Measuring the Distinctness of Neural Representations Using Multi-Voxel Spatial fMRI Patterns

Figure 1C illustrates the hypothesis tested in the present study: In people who can perceptually discriminate /r/ from /l/, the spatial patterns of fMRI activation evoked in their brains by those stimuli will be quite distinct, whereas in Japanese speakers, for whom /r/ and /l/ are difficult to tell apart, the evoked patterns of neural activity will be very similar. A crucial aspect of this hypothesis is illustrated in Figure 1C: The most distinct pair of fMRI activity patterns, in the rightmost column, is intermingled in the same part of cortex, and they contain equal overall amounts of activation. Thus, on this hypothesis, the relation between neural activity and levels of behavioral performance would be completely different from the "more activation is better" scenario sketched in Figure 1A. Instead, a person's ability to hear the difference between /r/ and /l/ would depend on how distinct those sounds' neural representations are from each other.

In standard fMRI analysis, spatial smoothing is applied to the blood oxygen level-dependent (BOLD) images. This smoothing has the effect of pooling the activations from voxels within a local spatial neighborhood, thereby improving the signal-to-noise ratio. However, as Figure 1C illustrates, stimuli that evoke quite distinct spatial fMRI patterns may produce equal total amounts of activation within a local spatial area. After spatial smoothing, the pattern differences will have been obliterated, and hence standard fMRI analysis will find the resulting smoothed average activations to be indistinguishable. It is for this reason that previous studies have concentrated mostly on comparing speech against nonspeech (Binder et al. 2000; Scott et al. 2000; Benson et al. 2001; Liebenthal et al. 2005) rather than comparing one speech token against another.

In the present study, we adopted an alternative technique: analyzing the unsmoothed local spatial fMRI patterns. Recent studies have shown that multi-voxel spatial patterns in fMRI data do indeed contain information not revealed by conventional statistical analyses (Haxby et al. 2001; Cox and Savoy 2003; Kamitani and Tong 2005; Polyn et al. 2005; Kriegeskorte et al. 2006; Hampton and O'Doherty 2007; Haynes et al. 2007; Pessoa and Padmala 2007; Serences and Boynton 2007; Williams et al. 2007), but the question of whether such patterns can be used to predict individual differences in behavioral ability has not been addressed.

A crucial advance in multi-voxel pattern analysis was made by Kriegeskorte et al. (2006), who proposed measuring the statistical information obtainable from within the local spatial neighborhood, or "sphere of information," at each point of the brain. This method, "information-based fMRI," allows a number to

be assigned to each voxel in the brain, where this number reflects not the amount of fMRI activation at that voxel but instead the statistical information contained in the local neighborhood centered on that voxel.

In the present study, we exploited this information-based approach to derive a quantitative measure of how distinct the neural representations of different syllabic stimuli were from each other. The specific measure that we used was the degree to which a classifier algorithm was able to separate the spatial fMRI patterns evoked by the stimuli. This neural measure allowed us to predict individual differences in the subjects' behavioral ability to tell the syllabic stimuli apart.

Native Language, Formant Space, and Discriminability

The sounds /r/ and /l/ are phonemic in English, meaning that they distinguish between words with different meanings, but in Japanese they are not (Tsujimura 2007). This phonological fact is borne out in psycholinguistic studies which demonstrate Japanese speakers' difficulty in perceiving the /r/-/l/ distinction (Goto 1971; Miyawaki et al. 1975; Iverson et al. 2003; Zhang et al. 2005). Consistent with the perceptual literature, this study manipulates 2 of the several spectral and temporal differences that distinguish these 2 sounds in American English, namely the frequencies of the second and third formants when other cues are held constant. (Formants are spectral prominences, or peaks in amplitude in the frequency spectrum, which vary mainly with the length and shape of the vocal tract.) Figure 2 shows how the sounds /r/ and /l/ are positioned in this F2/F3 formant space, for English and Japanese speakers. In particular, in English speakers /r/ and /l/ form 2 distinct phonemic categories, with /r/ corresponding to lower F3 values and /l/ to higher F3 values. Therefore, a change between high and low F3 produces a change in phonemic category and is hence easy for English speakers to perceive.

In contrast, Japanese speakers do not have distinct phonemic categories corresponding to high and low F3, as is shown in Figure 2B. For them, changes in F3 are a type of allophonic variation, meaning that such changes may alter how the stimulus sounds, but not in a way that changes the phonemic label, that is, word meaning should not be affected by such changes. Changes in the F2 dimension do not produce phonemic category changes for either English or Japanese subjects. Such changes are correspondingly hard to perceive for both groups, although Japanese speakers' sensitivity to F2 differences is slightly greater than it is for English speakers. The

perceptual sensitivity values, d' , of the English and Japanese speakers to F2 differences and F3 differences are shown in Supplementary Figure S1.

Also shown in Figure 2 are the 4 regions in formant space from which our fMRI stimuli were drawn, covering the 4 possible combinations of F2 and F3: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3, and Low-F2/Low-F3. This figure is a schematic simplification. In the Supplementary Material, a detailed plot is shown (Supplementary Fig. S2) of all the stimuli used in formant space, their F2 and F3 frequencies, the perceptual category ratings assigned to them by the English and Japanese speakers, and precisely which stimuli were used in the fMRI experiment. Spectrograms of the stimuli are shown in Supplementary Figure S3.

A Prediction: fMRI Pattern Separability Predicts Perceptual Discriminability

Given the above, the hypothesis that the statistical separability of fMRI patterns is related to perceptual discriminability leads to the following prediction: In English speakers, sounds that differ along the F3 dimension will evoke fMRI patterns that are more separable from each other than do sounds that differ in F2. Conversely, in Japanese speakers, F2 differences will lead to more separable fMRI patterns than will F3 differences. If the pattern separability tracks perceptual discriminability not only in its direction but also in its magnitude, then, in the English speakers, F3 differences will be a great deal more separable than F2 differences, but, in the Japanese, F2 differences will be just slightly more separable than F3 differences. In the following discussion, the term "separability" will be used to refer to the degree to which spatial fMRI patterns can be statistically distinguished from each other and the term "discriminability" will be used to describe people's abilities to tell stimuli apart perceptually.

Testing the Prediction

To test this hypothesis, we performed fMRI scans of 10 native American English and 10 native Japanese speakers while they listened to /ra/ and /la/ stimuli that differed along the F2 and F3 dimensions. The stimuli used were synthesized speech, and all ended with the constant vowel segment /a/, that is, /ra/ and /la/. The scans used a simple blocked design, with a single stimulus presented 12 times in each 25-s block, and the subjects pressing a button at the end of each block to identify the stimulus as either /ra/ or /la/. Outside of the scanner, the subjects were asked to make same/different judgments in an

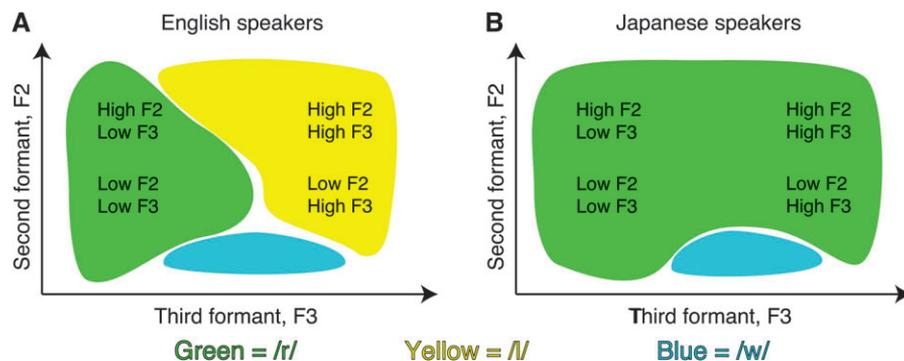


Figure 2. A simplified illustration of how different stimuli in F2/F3 formant space are perceived by English speakers (A) and Japanese speakers (B). The locations in this formant space of the 4 types of stimuli presented in the fMRI experiment are shown: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3, and Low-F2/Low-F3.

AX discrimination task, for the range of possible high/low F2/F3 stimulus pairings.

We then analyzed the resulting fMRI data to determine the separability of the fMRI patterns evoked by the various stimuli. The fMRI data were unsmoothed and were spatially normalized to the standard International Consortium for Brain Mapping 152 (ICBM152) echo-planar imaging (EPI) template brain without changing the original native-space voxel size. We used the sphere of information approach developed by Kriegeskorte et al. (2006), collecting the fMRI time courses in a 2-voxel-radius discretized sphere, with the sphere centered on each voxel in the brain in turn. Such a “sphere” contains 33 voxels, thereby generating a 33-element vector of fMRI activation values at each time point, with each vector corresponding to a particular spatial pattern of voxel activations in the sphere at a given moment in time.

To calculate the statistical separability of the fMRI patterns evoked by 2 different stimulus conditions, the 33-element activation vectors corresponding to the time points when those conditions were taking place were given as input to a linear support vector machine (SVM), with the desired output being the 2 condition labels. SVMs are algorithms that attempt to separate the data into classes, such that the margin between the decision boundary and the closest data points is as large as possible.

As our measure of pattern separability, we used the percentage correct achieved by the trained SVM, for each voxel’s Kriegeskorte sphere. Six-fold cross-validation was used (see Cross-Validation, Overfitting, and Noise below). The resulting “contrast image” consisted of the average percentage correct obtained by the SVM on the 6 cross-validation test sets, for the sphere of information centered on each corresponding voxel.

Observed percentage correct values on the test set generally varied between 50% (the chance level of performance), in areas such as white matter and the ventricles where no task-related activity was present, up to around 65–70% in some gray matter regions. Viewing the 2-category classification as a binomial random process with variance $np(1-p)$ where $p = 1/2$, the standard deviation (SD) of chance performance would be $\sqrt{100 \times 1/2 \times 1/2}$, which is equal to 5%. Thus, a test set score of 65% correct is 3 SDs above the mean.

Although using a nonlinear kernel in the SVM would have resulted in higher percentage correct scores, the advantage of a linear SVM is that it draws a simple planar decision boundary through the space of voxel activations. Nonlinear SVMs can create complicated curved decision boundaries, which are harder to interpret; it could become unclear whether the percentage correct tells us more about the statistics of the spatial patterns in the fMRI data or the mathematical agility of the nonlinear decision boundary.

Using the linear SVM, we calculated, for each subject, 2 different measures of statistical separability: the separability of fMRI patterns evoked by the high-F3 stimuli from those evoked by the low-F3 stimuli and the separability of high-F2 versus low-F2. This allowed us to calculate an F3-minus-F2 separability contrast image for each subject. These contrast images were passed into a second-level random-effects 2-sample *t*-test, looking for areas where the F3-minus-F2 separability was larger in the English speakers than the Japanese speakers.

Materials and Methods

fMRI Scanning

fMRI scans were carried out on a GE Signa 1.5-T scanner at the University of Washington. A standard BOLD EPI sequence was used:

time repetition (TR) = 2000 ms, time echo = 40 ms, field of view = 240 × 240 mm, 20 slices, voxels 3.75 × 3.75 mm, slice thickness = 4.5 mm, and interslice interval = 1.0 mm.

Each subject performed 2 functional runs, lasting 276 TRs each (552 s). During the scan, the subjects performed a syllable identification task, presented in a simple blocked design. In each block, a single syllable was presented once every 2 s, for 24 s in all. Each stimulus lasted 270 ms. At the end of each block, subjects had 5 s to press a button, indicating whether they perceived the syllable to be /ra/ or /la/. The auditory stimuli were presented using Avotec MRI-compatible headphones. After each task block, there was a 16-s rest block.

The question of whether to have the subjects respond after each stimulus within the block, as opposed to responding just once at the end of the block, hinges on whether one is more interested in neural activity related to the perception of the stimulus or related to the decision-making process underlying the responses. In the present study, we are more interested in the perception-related activity, so we chose to minimize the number of subject responses. Because all the stimuli within a block were the same, one response at the end of each block was sufficient.

There were 12 blocks in all per run, with 3 for each of the 4 different stimulus types: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3, and Low-F2/Low-F3. The different types of block were pseudorandomly ordered across the 2 functional runs.

MRI Data Processing

For tracing each individual subject’s right Heschl’s gyrus, the high-resolution (0.9375 × 0.9375 × 1.4 mm) T1 anatomical image was rigidly coregistered to their mean BOLD image and then spatially normalized using the parameters that were calculated from the BOLD normalization. The Heschl’s regions of interest (ROIs) were manually traced onto the underlying anatomy using FSLview (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain 2007). The mean of the 20 subjects’ coregistered and normalized anatomical images was used as the anatomical underlay for all the MRI cross-sections in the paper.

The BOLD images were preprocessed using Statistical Parametric Mapping, version 2 (SPM2) (Poline et al. 1997). They were motion corrected and then spatially normalized to the ICBM152 EPI template, preserving the original 3.75 × 3.75 × 5.5 mm voxel size. This normalization step was carried out before performing the information-based fMRI analysis, as it also is in standard general linear model (GLM) analyses, so that the spheres of information for all the subjects would encompass the same areas of neural tissue. Note that spatial normalization is independent from the grid structure of the slicing used to define voxels. By keeping this voxel grid size the same, we sought to minimize the interpolation and partial-volume-averaging effects described above.

For the information-based fMRI analysis (Kriegeskorte et al. 2006), no spatial smoothing was applied. Standard fMRI analysis was also carried out using SPM2. For that analysis, a separate copy of the data were spatially smoothed using a 6-mm full width half maximum Gaussian.

The voxels’ time courses were extracted and were high-pass filtered with a 300-s cutoff, in order to remove slow drifts. No low-pass temporal whitening filter was applied. After being zero-meaned, the time courses were then ready to be used for the information-based fMRI analysis.

Information-Based fMRI Analysis

For each voxel in the brain, the local spatial neighborhood of voxels was extracted, using a discrete sphere of radius equal to 2 voxels, making a sphere of information (Kriegeskorte et al. 2006) containing 33 voxels in all.

Two pairs of conditions were compared against each other: F3-high versus F3-low and F2-high versus F2-low. The time points corresponding to each condition were calculated by convolving the base condition time course by a haemodynamic response function and then picking those time points where the convolved result exceeded its mean value. The 33-element activation vectors corresponding to Conditions A and B, for each of the above pairings, were then passed into a linear SVM.

We used the Lagrangian SVM algorithm, developed by Mangasarian and Musicant (2000, 2001). Their Matlab code implementing the

algorithm is downloadable from <http://www.cs.wisc.edu/dmi/lsvm/>. The SVM was fitted to the data set using a maximum of 100 iterations. Six-fold cross-validation was used (see Cross-Validation, Overfitting, and Noise below). On a 2.5-GHz Intel Core 2 Duo CPU running Matlab 7.3 on Fedora Core 6 linux, each brain full of data took approximately 15 min to process per each cross-validation repetition. The resulting contrast image consisted of the average percentage correct obtained by the SVM on the 6 cross-validation test sets, for the sphere of information centered on each corresponding voxel.

For each subject, the F2 high-versus-low contrast image was subtracted from the F3 high-versus-low image to provide a measure of how much more separable the spatial patterns elicited by F3 differences were than F2 differences. These F3-minus-F2 contrast images were then passed into a second-level random-effects analysis in SPM2, using a 2-sample *t*-test to determine where the F3-minus-F2 value was higher in the English subjects than in the Japanese.

Note that the SVM did not have any information about the temporal order of the TRs within a condition; all the time points corresponding to one condition were lumped together into a single class without any time indexing. Other groups have incorporated temporal information into their classifiers (Mourao-Miranda et al. 2007), and this could be an interesting area of investigation in our future work. However, it was not done in the present study.

Cross-Validation, Overfitting, and Noise

Typically, when a classifier algorithm is trained on a data set, the purpose is so that the trained up classifier can subsequently be used to analyze new data. The trained classifier will perform poorly on the new data if it learned aspects of the training set that are not representative of the classes in the broader population. If that happens, then the classifier will fail to generalize from the training set to subsequent test sets. This is known as overfitting (Bishop 1995).

To guard against this, we performed 6-fold cross-validation, which involved the following: the 2 stimulus conditions (e.g., high-F3 and low-F3) had 12 blocks each, spread across the 2 runs. For each iteration of the cross-validation, 1 block from each condition was randomly selected (without replacement) to be used in the test set and the other 10 blocks were used for training. Thus, over the course of the 6 iterations, every stimulus block participated in the test set exactly once. Because the blocks were separated from each other by 16 s of rest, any potential hemodynamically induced temporal correlation between the training set and test set fMRI data points was prevented. The mean percentage correct obtained across these 6 test sets was then calculated to give the final output.

Although cross-validation brings with it the crucial benefit of avoiding overfitting, it also carries costs. The first cost is computational:

to compute 6-fold cross-validation takes 6 times as long. The second cost is that it adds noise to the analyses: ideally an analysis would use all the available data points, but each partitioning of the data into training and test sets prevents some arbitrary subset of the data points from being modeled. The minimally disruptive partitioning is “leave one out” cross-validation, in which only one data point is used for the test set and every data point is used in turn. However, this also carries the highest computational cost, taking *n* times longer to calculate for *n* data points. The 6-fold cross-validation used here occupies a middle ground, but it should be borne in mind that the statistical significance of the results presented here is slightly lower than it would be if lengthy leave one out cross-validation had been used.

Nonparametric Multiple Comparisons Correction

Because the analysis presented here uses unsmoothed data, the theory of Gaussian random fields (Poline et al. 1997) does not apply. Thus, no corrected *P* value based on the effective number of spatially independent tests is currently available. However, multiple comparisons correction can also be performed using nonparametric methods, which do not rely on any assumptions about the spatial smoothness of the data (Nichols and Holmes 2002) (we would like to thank Russ Poldrack for suggesting the use of a nonparametric approach here). Figure 3A shows the right Heschl’s ROI thresholded at the uncorrected *P* value of 0.001. Supplementary Figure S4 shows that this is significant at the family-wise error-corrected *P* value of *P* = 0.083, as calculated using the statistical nonparametric mapping (SnPM) nonparametric analysis package (Nichols and Holmes 2002).

As is noted in the section Cross-Validation, Overfitting, and Noise, the *P* values obtained from cross-validated classifier analyses are likely to be somewhat overconservative, as the partitioning of the data into training and test sets introduces a source of noise that is not present in the data itself. Ideally, a process for generating corrected *P* values would take such factors into account and would modify the resultant *P* values accordingly. No such method is currently available.

In the absence of such a method, a nonparametric multiple comparisons approach lacking any classifier-related information is probably the best available approach, albeit perhaps an overly conservative one. SnPM randomly permutes the labels of Japanese and English across the 20 subjects, calculates the *T* value of the difference between the “English” and “Japanese” speakers for each such relabeling, and then compares the *T* value from the true subject labeling to the distribution of *T* values from the many random labelings. For each relabeling, a single summary statistic is calculated, which in this case is the size of the biggest above-threshold cluster that emerges from that labeling’s English versus Japanese group difference. The distribution of these maximum cluster sizes can also be

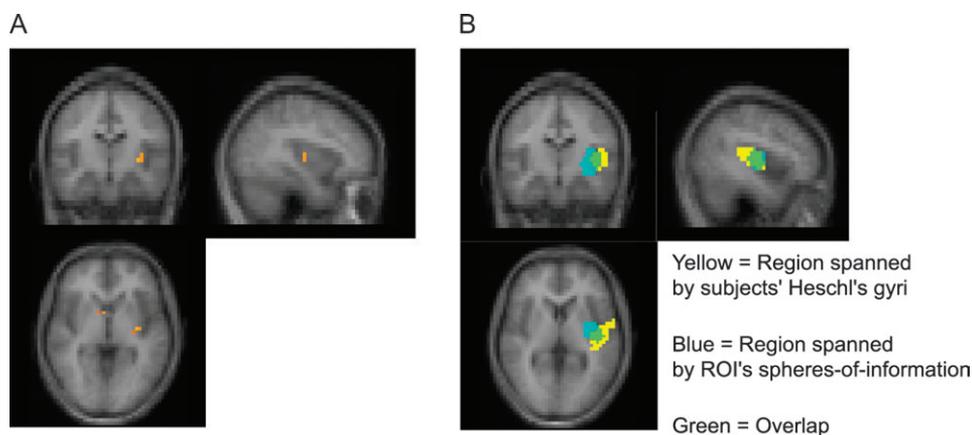


Figure 3. (A) Areas in which the F3-minus-F2 fMRI pattern separability is greater for the English speakers than for the Japanese. It can be seen that this random-effects contrast reaches its maximal significance in right Heschl’s gyrus (primary auditory cortex). Note that this is not “activation,” as in conventional fMRI analysis, but is instead the amount of information present at each voxel (Kriegeskorte et al. 2006). The anatomical underlay image is the mean of the 20 subjects’ coregistered and normalized anatomical scans. (B) The local spatial patterns from which the statistics in panel (A) are calculated extend beyond the voxels shown there. The region spanned by the voxels that contribute to that ROI are shown in yellow. Shown in blue are the voxels spanned by the Heschl’s gyri of the subjects, manually traced on each individual’s *T*₁ anatomical scan. The areas of overlap are shown in green.

compared against the maximum cluster size that arises from the true subject labeling. Because the “English” and “Japanese” labels are applied to entire brain at once, rather than voxel by voxel, and because the maximum cluster size value is just one single number for the whole brain, the *P* value describing how the true labeling’s maximum cluster size compares to the others is the result of just a single statistical comparison, not of multiple comparisons.

Speech Stimuli

The speech stimuli were synthesized using the Hlsyn program (Hanson and Stevens 2002), from Sensimetrics Corporation, Cambridge, MA. This program allows speech synthesis parameters to be entered at a high level (HL parameters) or at the lower level of the Klatt formant synthesizer (KL parameters). In the present study, all values were entered as KL parameters.

At the first stage, a broad array of different stimuli was generated, differing in the starting frequencies of their F2 and F3 formant transitions. There were 57 such tokens, varying in F2 (701–2201 Hz, in 300-Hz steps) and F3 (725–3925 Hz, in 300-Hz steps). These stimuli were then presented to American English ($n = 9$) and Japanese ($n = 14$) native speakers, who were asked to give their phonetic identification and goodness ratings. Note that these were not the subjects who participated in the fMRI experiment but instead were a totally nonoverlapping group. Each syllable was identified and was rated for goodness on a 7-point scale (1 = very poor, 7 = very good).

The details of the stimulus array and the behavioral results are shown in Supplementary Figure S2. The English speakers identified the syllables as English /ra/, /la/, and /wa/. The Japanese speakers identified the syllables as Japanese /ra/ and /wa/. The stimuli that received the highest rating scores are marked in Supplementary Figure S2 with red squares.

Twelve of the stimuli that received the highest perceptual goodness ratings were selected from the array to be used for the subsequent fMRI experiment with different subjects. These stimuli were of 4 sorts: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3, and Low-F2/Low-F3. The selected stimuli are highlighted in Supplementary Figure S2 by the black outlined regions.

The detailed stimulus synthesis parameters are as follows: syllable duration = 270 ms, consonant duration = 60 ms, consonant-to-vowel F2/F3 transition duration = 30 ms, vowel /a/ duration = 180 ms.

Consonant formants: F1 = 365 Hz, bandwidth = 200 Hz; F2 bandwidth = 100 Hz; F3 bandwidth = 150 Hz; F4 = 4512 Hz, bandwidth = 100 Hz.

Vowel formants: F1 = 965 Hz, bandwidth = 200 Hz; F2 = 1807 Hz, bandwidth = 100 Hz; F3 = 3164 Hz, bandwidth = 150 Hz; F4 = 4512 Hz, bandwidth = 400 Hz.

F0: 221 Hz in the consonant portion; 247 Hz in the vowel portion.

Tables providing the complete details of all the speech synthesis parameters can be found in the Supplementary Material (Table S1 on p. 24, Table S2 on p. 25, and Table S3 on p. 26).

Behavioral Testing Outside the Scanner

Outside of the scanner, the subjects were presented with an AX same/different discrimination task in order to determine their d' sensitivity scores to changes along the F2 and F3 dimensions. Pairs of stimuli from the High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3, and Low-F2/Low-F3 regions of the formant space were presented: 96 trials in total, of which 24 were same-pair trials and 72 were different pairs. The stimuli were pseudorandomly ordered for each subject, and the within-pair presentation ordering was counterbalanced. The onset-to-onset interstimulus interval for each pair was 800 ms. For signal detection theory analysis, the pairings were grouped according to whether the stimuli within the pair differed in F2 or F3. Hit and false alarm rates were calculated and then converted into d' scores.

Subjects

Twenty subjects participated in the experiment: 10 were native American English speakers and 10 were native Japanese speakers. The Japanese subjects had lived in the United States for between 6 and 12 months. The English speakers were aged between 21 and 34 years

and were all female. The Japanese speakers were aged between 20 and 37 years; 4 were male. All subjects gave written informed consent, as approved by the Human Subjects Institutional Review Board of the University of Washington.

Results

English versus Japanese Differences in Pattern Separability in Right Heschl’s Gyrus

The results of this group-level analysis are shown in Figure 3A. It can be seen that one region is primarily “activated”: right Heschl’s gyrus (primary auditory cortex). The word “activated” is in quotation marks here because the quantity being plotted is not fMRI BOLD signal change, as in a conventional analysis, but instead is the amount of information present at each voxel (Kriegeskorte et al. 2006), in this instance the degree of separability of the various evoked activation patterns.

Another sense in which the “activation” shown in Figure 3A must be interpreted differently from standard fMRI results is that it shows the locations of the centers of the spheres of information, in which the separability values of the local spatial fMRI patterns are calculated. The voxels within each such spatial neighborhood extend up to 14 mm away from each sphere’s center, and all these voxels can contribute equally to the pattern separability statistic that is assigned to the center voxel. Although standard fMRI analyses often resample BOLD images into smaller voxels, the information-based analyses here are calculated on voxels that have been neither smoothed nor resized as such operations could disrupt spatial patterns in the data. Figure 3B shows, in blue, the region spanned by the voxels that contribute to the ROI in panel (A). Shown in yellow are the voxels spanned by the Heschl’s gyri of the subjects, manually traced on each individual’s T1 anatomical scan. The areas of overlap are shown in green. It can be seen that because the spheres of information are centered on the anterior medial part of Heschl’s gyrus, they include some voxels from the neighboring posterior insula, in addition to those from auditory cortex. The local fMRI pattern separability is a collective property of all of a sphere’s voxels taken together and cannot be pinpointed to the sphere’s center, despite the graphical convenience of representing it in that way.

In addition to looking at brain-wide statistical maps, the same neural data can be examined in a close-up view, by plotting the individual subjects’ data drawn from specific ROIs. Figure 4A shows that the statistical separability of the fMRI patterns in right auditory cortex strongly predicts how well the subjects could perceptually discriminate the stimuli ($r = 0.796$, $P < 3 \times 10^{-5}$). It is important to emphasize that the right auditory cortex ROI emerged as the result of a purely group-wise comparison of pattern separability: English versus Japanese. It would be perfectly possible, given the between-group pattern separability differences, for the within-group differences to fail to correlate with the perceptual d' scores. For example, the within-group patterns of the Japanese and English subjects in the scatterplot could be flipped left to right without altering the group-wise difference in the slightest. Such a pattern, however, would greatly reduce any brain-behavior correlation sought across all 20 subjects. It is therefore noteworthy that the right auditory cortex reveals such a strong correlation across individuals as well as across groups. This can be quantified by performing a partial correlation ($r = 0.389$, $P < 0.05$, 1-tailed), which shows

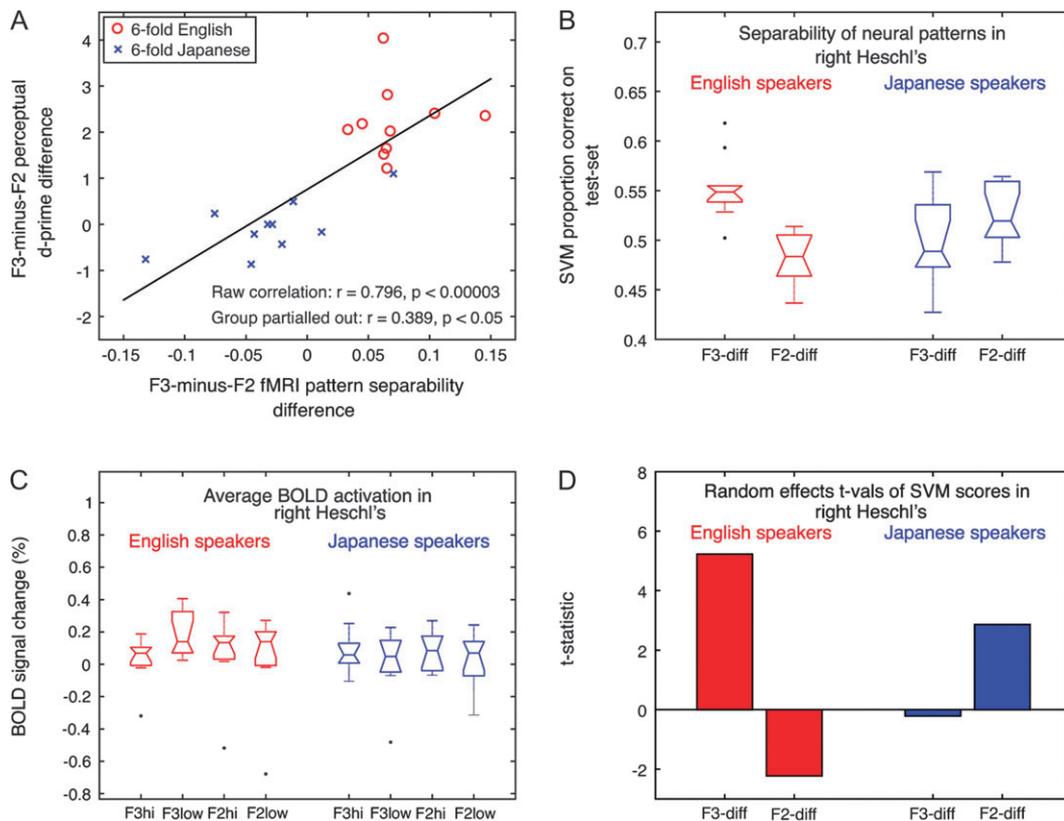


Figure 4. (A) The statistical separability of the fMRI patterns in right auditory cortex strongly predicts how well the subjects could perceptually discriminate the stimuli. This correlation holds across individuals as well as groups, as shown by the correlation with group membership partialled out. (B) The group-wise differences in the separability of the fMRI patterns for the English and Japanese speakers, showing the F3 differences and F2 differences in their own right, rather than the compound F3-minus-F2 difference that is plotted along the x-axis of panel (A). (C) Despite the separability of the evoked spatial patterns, the average neural activation in this region, measured by conventional fMRI analysis, does not differ significantly along either of the formant dimensions. All boxplots follow the standard conventions: the box center is the median, the ends are the interquartile ranges, the whiskers are 1.5 times the interquartile ranges, dots are outliers beyond the whisker range, and the notches are such that boxplots whose notches do not overlap have different medians at the 5% significance level.

that the positive correlation which was observed to hold across individuals remains significantly greater than zero, even after removing any effects of group membership.

As well as the right auditory cortex information-bearing activation, the English versus Japanese pattern separability contrast also revealed some activation in the head of the caudate, bilaterally. This activation was not as statistically significant as the right auditory cluster, nor did it show that cluster's correlation across individuals as well as groups. Nonetheless, this finding is of some interest in the light of a number of recent studies showing the involvement of the caudate in bilingual language control (Crinion et al. 2006; Wang et al. 2007; Abutalebi et al. 2008; Ketteler et al. 2008). The full data for left and right caudate, respectively, are plotted in Supplementary Figures S10–S12.

Differences in fMRI Pattern Separability without Differences in Average Local Activation

The group-wise differences in the separability of the fMRI patterns are shown in Figure 4B, and the average BOLD activations for the F3-high, F3-low, F2-high, and F2-low conditions for the 2 groups are shown in Figure 4C. Despite the separability of the evoked spatial patterns, the average neural activation in this region, which is what conventional fMRI analyses measure, did not differ significantly along either of the formant dimensions. It can be seen that all the different speech stimuli activated auditory cortex

to approximately the same degree. However, they created subtly different spatial patterns of activation that the present analysis method can detect but to which conventional fMRI analyses are blind.

The fact that auditory cortex responds with more or less equal intensity to different syllables is neither surprising nor noteworthy in itself. The point that we wish to emphasize here is that even despite this lack of any local average activation intensity, robust pattern differences can still occur. This is precisely the situation illustrated schematically in Figure 1.

The Converse: BOLD Activation without Between-Condition Differences in fMRI Pattern Separability

Whereas the right primary auditory cortex showed differences in pattern separability without corresponding changes in average local activation, other regions of the brain exhibited the opposite effect. That is, these regions are activated by the speech stimuli and are hence revealed by a standard fMRI speech-versus-nonspeech contrast, but their spatial patterns of fMRI activation across the different formant conditions are not separable from each other.

Using a conventional GLM fMRI analysis of the data, and looking for regions activated on average by all the speech stimuli, the standard set of speech-related regions were found: Broca's area on the left and the superior temporal gyrus

bilaterally, consistent with other fMRI studies of /r/ and /l/ perception (Callan et al. 2003; Jacquemot et al. 2003).

Classical Speech Areas and Pattern Separability

The regions found to be activated by the speech stimuli in our conventional fMRI analysis and the details of their responses to the different F2 and F3 conditions are shown in Supplementary Figures S5, S6, S7, and S8, respectively. As well as the superior temporal gyrus (STG) and Broca's regions, Supplementary Figure S9 plots the results for a left Heschl's gyrus ROI, constructed as a mirror image of the right Heschl's ROI derived from the pattern separability random-effects analysis. This allows a direct comparison of the behavior of the primary auditory cortices in the 2 hemispheres.

These figures show that these classical language areas are strongly activated by all the speech conditions but without significant activation differences between the different conditions. Moreover, the different conditions were not separable by the SVM classifier in these regions, unlike in right primary auditory cortex. Of these classical speech regions, Broca's area comes the closest to being able to distinguish between any of the conditions, showing higher BOLD activation for F2-low than F2-high stimuli in the Japanese speakers and a tendency, albeit nonsignificant, for greater separability of F3 differences than F2 differences in the English speakers. It is possible that more refined statistical pattern recognition analyses that are presented here might reveal more information being carried in Broca's area.

For a better overview of the pattern separability of the fMRI activation elicited by the various speech stimuli, it is necessary to examine not just the activation in particular ROIs but also the information maps across the whole brain, viewed at multiple slices and at multiple statistical thresholds. These are shown in Supplementary Figure S13, which shows F3 and F2 separability individually, for the Japanese and English speakers, and Supplementary Figure S14, which shows statistical maps of the various F3 versus F2 and English versus Japanese comparisons.

For a full discussion of possible reasons why information-bearing activation was found more in the right hemisphere than in the left, see sections *Why Does Phonetic Information Show Up in Right Auditory Cortex, Rather Than the Left?* and *Why Is Signal Absent from the Left, when Phonetic Processing Is Undoubtedly Taking Place There?* in the Discussion below.

Discussion

The results presented here demonstrate for the first time that it is possible to use spatial patterns of brain activation to predict individual differences in people's perceptual ability. Unlike many previous studies, in which increased activation corresponds to better behavioral performance, greater perceptual ability in the present case stems from the underlying neural representations being more distinguishable from each other. As Figure 4C shows, the average BOLD activation evoked by /ra/ and /la/ in English speakers is no greater than it is in Japanese speakers. However, the statistical separability of the spatial fMRI patterns is greater in English speakers (Fig. 4B), and this increased neural separability predicts individual differences in people's perceptual ability to tell /ra/ and /la/ apart (Fig. 4A).

This finding was made possible by using the information-based fMRI approach (Kriegeskorte et al. 2006) to access different neural representations that were colocalized within the same cortical area, then by quantifying the distinctness of

those representations using a linear classifier algorithm, and finally by comparing that derived neural measure to the subjects' behavioral discrimination scores.

This shows not only that spatial fMRI patterns can be directly connected to behavior but moreover that the statistical properties of these fMRI patterns can quantitatively predict very specific aspects of behavioral performance, both across groups and across individuals. It should be noted that the fMRI pattern separability measure was calculated purely as a function of the neural data and did not incorporate any information about subjects' group membership or behavior.

Relation to Previous Studies

There has been very little evidence to date that fMRI spatial patterns have any relation to behavior at all, as opposed to merely being epiphenomena. Certainly these fMRI patterns arise, albeit indirectly, from neural activation, and neural activation is what gives rise to all perception and cognition; the patterns would therefore be expected to be correlated with ongoing perceptual and cognitive events, as has indeed been shown to be the case (Haynes and Rees 2005; Polyn et al. 2005; Haynes et al. 2007; Williams et al. 2007). The present study demonstrates that neural data alone, without incorporating any information about behavioral responses, can indeed be used to predict behavior, and more specifically that the distinctness of a given individual's neural representations can predict their level of perceptual ability.

Our finding that the distinctness of neural phonemic representations varies across individuals raises the question of how such differences arise. Studies in infants have shown that differences in the language environments experienced in the first few months of life play a crucial role in shaping how formant space becomes divided into distinct phonemic categories. The theory of "neural commitment" (Kuhl 2004) proposes that after those native language categories have developed by tracking the statistics of the early language environment, it becomes much harder for different language environments experienced later in life to reshape auditory cortex. Even though cortex may still be plastic (Buonomano and Merzenich 1998), its acquired tuning makes it less sensitive to the statistics of any new and different environments that might try to restructure it. Studies in adults have shown that the native language environment in which a speaker was raised is reflected in the set of brain regions that are activated (Callan et al. 2003; Jacquemot et al. 2003), the perceptual structure of formant space (Iverson et al. 2003), and the amount of neural tissue recruited by phonemic processing (Zhang et al. 2005). The present results move beyond such studies by showing how the distinctness of the brain's neural representations of language can be directly quantified.

Why Does Phonetic Information Show Up in Right Auditory Cortex, Rather Than the Left?

Although speech stimuli generally activate both left and right auditory cortex bilaterally, language-specific contrasts such as speech-versus-nonspeech almost always reveal greater activation on the left. Given that, it might appear somewhat surprising that in the present study it was right auditory cortex, rather than the left, whose local spatial fMRI patterns were found to contain the greatest amount of information about which speech stimulus gave rise to them.

However, several recent studies have found predominantly right-hemisphere speech-related activity. In 2 separate studies, Alain et al. (2007) and Alain and Snyder (2008) presented their subjects with the task of identifying 2 vowels that were played simultaneously. The subjects found this task to be very difficult at first, but then learned the necessary discrimination as their performance improved during the first hour of testing. Simultaneous event-related potential recordings revealed that this improvement in performance was paralleled by enhancements in an early evoked response (around 130 ms) localized in the right auditory cortex. Using a different approach, Kujala et al. (2002) studied mismatch negativity (MMN) responses to consonant contrasts which were either embedded in the middle of a word or presented in isolation. They found that MMN responses were larger for the embedded contrasts in right auditory cortex but not in the left.

Two recent studies, using magnetoencephalography (Luo and Poeppel 2007) and electroencephalography (Abrams et al. 2008), respectively, have found evidence that the right auditory cortex is involved in processing some of the slower aspects of the speech signal, in a time window on the order of around 200 ms. This is the timescale in which syllables occur. Luo and Poeppel (2007) found that the phase pattern of theta band (4–8 Hz) responses in right auditory cortex carried information about presented speech, and Abrams et al. (2008) produced evidence that right-hemisphere responses on a similar timescale tracked the overall acoustic envelope of the speech stimulus.

Most directly relevant to the present work is the recent fMRI study by Obleser et al. (2008), who showed that right auditory cortex is sensitive to the amount of spectral information present in the speech signal. Note that the spectral information in that study was a function of the formants and was intrinsically part of the intelligibility of the speech signal itself. Moreover, the right-hemisphere processing observed in that study was very close to the location found in the present work, namely in right Heschl's gyrus (see especially Figs 4A and 5A of Obleser et al. [2008]).

Thus, a growing body of recent experimental data now corroborates the finding that right auditory cortex is involved in speech processing. A possible explanation for why the right-hemisphere effects found in the above studies are not more commonly observed is that they all probed some of the less-often investigated aspect of speech perception, such as speech processing in the presence of auditory distractors (Kujala et al. 2002; Alain et al. 2007; Alain and Snyder 2008), the temporal envelope and syllabic structure of speech (Luo and Poeppel 2007; Abrams et al. 2008), or spectral degradation manipulated independently of temporal degradation (Obleser et al. 2008).

If the right hemisphere is indeed involved in speech processing, but in ways that are missed by standard experimental approaches, then it is worth asking which kinds of speech-related neural activation might be expected to be missed by a standard fMRI analysis. As Figure 1C illustrates, one scenario that may very easily occur is when different speech stimuli produce activation of equal local average intensity but with distinct local spatial patterns. Because a standard fMRI analysis measures only the local average intensity of activation, the pattern changes illustrated in Figure 1C would be completely invisible to it.

Moving from the hypothetical schemas of Figure 1 to the actual data plotted in Figure 4, it can be seen that this situation is exactly what was observed. The boxplots of right auditory cortex activation in Figure 4B show that the various F3 and F2

stimulus conditions differ greatly from each other in terms of how separable the spatial fMRI patterns that they elicit are. However, as Figure 4C shows, when we look not at the separability of spatial patterns but instead at average local activation, all the different stimulus conditions appear to be much the same. All the stimuli, whether /ra/ or /la/, induce some activation in auditory cortex and indeed they all induce approximately the same overall amount. However, they distribute that activation over subtly different spatial patterns. These differences can be captured by a pattern-based analysis, but to a standard fMRI analysis that only measures overall activation they must remain unseen.

Collectively, these considerations suggest the following. Left auditory cortex carries out a large amount of language-related processing, with the result that speech stimuli tend to activate it with greater overall intensity than on the right. However, right auditory cortex is also engaged in language processing but in ways that standard experimental approaches may easily miss. The data from the present study suggest that the overall level of speech-related activation in right auditory cortex does not become more intense; instead, its neural representations of different phonemes produce different spatial patterns of activation, with the local average staying unchanged. Previous fMRI studies have measured this unchanging average activation and have drawn the perhaps unwarranted conclusion that language processing must be happening elsewhere.

Why Is Signal Absent from the Left, When Phonetic Processing Is Undoubtedly Taking Place There?

Just as it is somewhat unusual to find evidence of phonetic processing in right auditory cortex, it is curious that similar signals do not also turn up on the left. We are certainly not claiming that left-hemisphere auditory areas are insensitive to the /r/-/l/ distinction. Such a claim would be unwarranted and almost certainly false. Moreover, the MRI signal itself on the left is robust, as shown by the standard fMRI analyses of speech-versus-rest in Supplementary Figure S5, which show the classical pattern of activation: bilateral but stronger on the left, especially in left STG and Broca's area. Why then the absence of fMRI pattern separability in left auditory areas?

A plausible hypothesis is that the phonetic representations in left auditory cortical areas may be spatially overlapping and intermingled with each other at a finer spatial scale than they are in the right hemisphere. In particular, if these overlapping representations are colocalized and fully contained within individual voxels, then multi-voxel pattern-based analyses will be unable to see them. By their very nature, these analyses seek activation patterns that are spatially distributed across several voxels. In contrast, a different method is best suited to finding overlapping representations contained within individual voxels, namely adaptation-fMRI (Grill-Spector and Malach 2001). Although standard fMRI is unable to distinguish between intravoxel overlapping activations, the method of adaptation-fMRI is designed to do just that.

A previous study from our group using adaptation-fMRI produced evidence that left-hemisphere speech areas do indeed contain intravoxel overlapping phonetic representations. Raizada and Poldrack (2007) studied the degree to which each voxel in the brain was sensitive to the phonetic category distinction between /ba/ and /da/ and found that the left supramarginal gyrus showed adaptation effects which tracked the category boundary.

This indicates that the left supramarginal gyrus contains voxels within which distinct but spatially intermingled populations of /ba/-sensitive and /da/-sensitive neurons coexist.

Figures illustrating direct left-versus-right comparisons for the right auditory cortex area revealed by our analyses and its mirror image equivalent ROI on the left are presented in Supplementary Figures S15–S17. These analyses show that the right auditory cortex produces strong and significant multi-voxel pattern effects, whereas left auditory cortex does not reach significance. (As remarked above, this is likely due to the left-hemisphere processing occurring at an intravoxel rather than a multi-voxel spatial scale). Given that we are not seeking to argue that left auditory cortex lacks sensitivity to phonemes, we would not expect the right-minus-left comparisons to reach significance and indeed they do not.

Taken together, the Raizada and Poldrack (2007) study and the present manuscript suggest the following hypothesis: left-hemisphere auditory cortical regions process phonetic category distinctions at an intravoxel spatial scale and that right-hemisphere regions perform such processing at a multi-voxel spatial scale. This naturally raises the question of whether adaptation-fMRI and multi-voxel pattern-based fMRI can be combined to investigate both spatial scales simultaneously. The “continuous carryover” design developed by Aguirre (2007) does just that, and indeed, we plan to conduct a study using such a design to investigate phonetic processing in future work.

Two very interesting neurophysiological studies are also worth mentioning, with reference to whether speech-related information might be encoded in neurons outside of what are classically thought of as “speech areas.” Recording in the primary auditory cortex of the ferret, bilaterally, while presenting phonemic stimuli, Mesgarani et al. (2008) found that the neural responses from a population of 90 neurons were sufficient to allow a linear SVM to decode which stimuli had been played. Their study is therefore quite analogous to the work presented here, except that they operated at the level of single neurons rather than fMRI voxels. The fact that primary auditory cortex was found to contain phonemic information both in their experiment and in the present study is especially striking. In an even more recent study, Remedios et al. (2009) recorded bilaterally from posterior insula cortex in the macaque and found neurons there which responded preferentially to vocal communication sounds. This posterior insula region directly neighbors primary auditory cortex and may, speculatively, be a monkey homolog of the posterior insula voxels shown in Figure 3B, which fell within the span of the spheres of information centered on the Heschl’s gyrus ROI.

No single study on its own can confirm whether or not the above hypotheses hold true. It will be interesting to see whether future studies using pattern-based analyses also find information to be present in auditory cortex on the right. Whether they do or do not, they will be measuring a quite different type of signal than the average local activation studied by standard fMRI. From these new types of measurements, one may expect new types of results.

Will Similar Results Be Found in Domains beyond Speech Perception?

As with any fMRI method, information-based fMRI processes only a spatially coarse picture of the sluggish hemodynamic events triggered by the underlying neural activity, which is of

course the process that really matters. Certainly, the linear decision boundary drawn by the SVM in the 33-dimensional space of a sphere of voxels bears little or no resemblance to the subtle neural mechanisms by which the brain extracts information. However, what this statistical artifice does show is that even the indirect MRI trace contains enough remnants of the true neural information processing that it can be meaningfully related to behavior. The present study demonstrates not only that such information is present, but also that this neural information is strongly predictive of individual’s perceptual abilities.

Beyond the domain of speech perception, there are a great many types of task in which behavioral success depends not only upon the degree to which a particular cognitive process is occurring but also upon the properties of the neural representations that are involved. However, the goal of quantifying the properties of such representations has proven elusive, as it is only in low-level sensory cortices that well-defined representational maps are available (Duncan and Boynton 2003, 2007). The study presented here lays out a path for possibly approaching that goal in higher-level cortical areas as well.

It is an open question whether similar results will be found to hold in other cognitive and perceptual processes and other brain regions. The existence of colocalized but distinct neural representations and processes is an extremely general phenomenon in the brain, and it is precisely these circumstances in which standard fMRI analyses face difficulties. Other fMRI methods, notably adaptation-fMRI (Grill-Spector and Malach 2001; Raizada and Poldrack 2007), have also attempted to tackle the problem of pulling apart distinct but colocalized representations. However, adaptation-fMRI techniques require more complex experimental designs and rest on larger sets of assumptions, which are testable in very specific domains (Boynton and Finney 2003) but whose broader validity is less certain. Therefore, there is a need for new approaches to studying neural representations, and multi-voxel pattern-based fMRI analyses may offer a useful set of tools for developing them. One example from a completely different task domain, indicating the potential generality of such methods, is an innovative recent study by Haynes et al. (2007) who found different local spatial patterns elicited by addition and subtraction tasks. The present study shows for the first time that the statistical separability of such patterns can predict individual differences in behavioral success. An analogous study in the domain of arithmetic would ask not only whether the patterns can predict whether an addition task or a subtraction task is being performed but more specifically whether the patterns can predict if the person is carrying out the arithmetic easily or with difficulty.

Possible Practical Applications and Testable Predictions

Given this new ability to probe neural representations that would otherwise be inaccessible, and to relate these representations directly to behavioral success, it is reasonable to ask what this might be useful for, over and above its interest as basic science. One concrete potential use might be for diagnosing whether behavioral difficulties are due to problems of competence or of performance. For example, 2 children may both achieve the same low score on a speech perception test. One possibility is that the representational distinctions necessary for

carrying out the test may not be properly formed in a child's brain. Alternatively, the underlying representational machinery, that is, the competence, may be functioning well, but performance issues such as attention or motivation may give rise to the low score. From the outside, measured behaviorally, these 2 alternatives are hard to pull apart, as the children's scores on the test are the same. However, by investigating the separability of the children's neural representations, measurable differences may emerge. A problem of representational competence would require a very different type of training-based intervention than would one of the performance. The link between such neural separability and behavior suggests an additional testable hypothesis: over the course of behavioral training, increases in performance and increases in the separability of the neural representations should go hand in hand. The questions of whether such hypotheses will turn out to hold true, and whether the new results presented here will generalize across other perceptual and cognitive processes, are likely to be fruitful topics for future investigation.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>

Funding

National Science Foundation (NSF) Science of Learning Center (NSF 0354453) to the University of Washington Learning in Informal and Formal Environments (LIFE) Center.

Notes

Our thanks to Geoff Boynton, Tanzeem Choudhury, Katherine Demuth, Ione Fine, Nikolaus Kriegeskorte, and Russ Poldrack for comments on the manuscript. Author contributions: R.D.S.R. conceived and developed the theoretical ideas for relating information-based fMRI to behavioral performance, wrote, and executed the computational analyses using Matlab scripts and SPM2, traced the individual subjects' anatomical ROIs, and was the primary author of the manuscript. F.M.T. designed and executed the fMRI scans and behavioral experiments, created the synthetic speech stimuli, analyzed the behavioral data, and conducted GLM fMRI analysis using BrainVoyager. H.M.L. assisted with the behavioral testing, fMRI scans, and GLM fMRI analysis. P.K.K. is the laboratory head, developed the theory of neural commitment, and provided overall scientific guidance. *Conflict of Interest:* None declared.

Address correspondence to Dr Feng-Ming Tsao. Email: tsaosph@ntu.edu.tw.

References

Abrams DA, Nicol T, Zecker S, Kraus N. 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci*. 28:3958–3965.

Abutalebi J, Annoni JM, Zimine I, Pegna AJ, Seghier ML, Lee-Jahnke H, Lazeyras F, Cappa SF, Khateb A. 2008. Language control and lexical competition in bilinguals: an event-related fMRI study. *Cereb Cortex*. 18:1496–1505.

Aguirre GK. 2007. Continuous carry-over designs for fMRI. *Neuroimage*. 35:1480–1494.

Alain C, Snyder JS. 2008. Age-related differences in auditory evoked responses during rapid perceptual learning. *Clin Neurophysiol*. 119:356–366.

Alain C, Snyder JS, He Y, Reinke KS. 2007. Changes in auditory cortex parallel rapid perceptual learning. *Cereb Cortex*. 17:1074–1084.

Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, Liberman AM. 2001. Parametrically dissociating speech and non-speech perception in the brain using fMRI. *Brain Lang*. 78:364–396.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*. 10:512–528.

Bishop CM. 1995. *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Boynton GM, Finney EM. 2003. Orientation-specific adaptation in human visual cortex. *J Neurosci*. 23:8781–8787.

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JD. 1998. Making memories: brain activity that predicts how well visual experience will be remembered. *Science*. 281:1185–1187.

Buonomano DV, Merzenich MM. 1998. Cortical plasticity: from synapses to maps. *Annu Rev Neurosci*. 21:149–186.

Callan D, Jones J, Callan A, 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.

Callan D, Tajima K, Callan A, Kubo R, Masaki S, Akahane-Yamada R. 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*. 19:113–124.

Cox D, Savoy R. 2003. Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*. 19:261–270.

Crinion J, Price CJ. 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*. 128:2858–2871.

Crinion J, Turner R, Grogan A, Hanakawa T, Noppeney U, Devlin JT, Aso T, Urayama S, Fukuyama H, Stockton K, et al. 2006. Language control in the bilingual brain. *Science*. 312:1537–1540.

Demb JB, Boynton GM, Heeger DJ. 1997. Brain activity in visual cortex predicts individual differences in reading performance. *Proc Natl Acad Sci USA*. 94:13363–13366.

Duncan R, Boynton G. 2003. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*. 38:659–671.

Duncan R, Boynton G. 2007. Tactile hyperacuity thresholds correlate with finger maps in primary somatosensory cortex (S1). *Cereb Cortex*. 17:2878–2891.

Gauthier I, Skudlarski P, Gore J, Anderson A. 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*. 3:191–197.

Goto H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*. 9:317–323.

Grabner R, Ansari D, Reishofer G, Stern E, Ebner F, Neuper C. 2007. Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage*. 38:346–356.

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.

Guion SG, Flege JE, Akahane-Yamada R, Pruitt JC. 2000. An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *J Acoust Soc Am*. 107:2711–2724.

Hampton AN, O'Doherty JP. 2007. Decoding the neural substrates of reward-related decision making with functional MRI. *Proc Natl Acad Sci USA*. 104:1377–1382.

Hanson HM, Stevens KN. 2002. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *J Acoust Soc Am*. 112:1158–1182.

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.

Haynes JD, Rees G. 2005. Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol*. 15:1301–1307.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading hidden intentions in the human brain. *Curr Biol*. 17:323–328.

- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, Siebert C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 87: B47-B57.
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E. 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci*. 23:9541-9546.
- Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*. 8:679-685.
- Ketteler D, Kastrau F, Vohn R, Huber W. 2008. The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *Neuroimage*. 39:2002-2009.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 103:3863-3868.
- Kuhl PK. 2004. Early language acquisition: cracking the speech code. *Nat Rev Neurosci*. 5:831-843.
- Kujala A, Alho K, Valle S, Sivonen P, Ilmoniemi RJ, Alku P, Naatanen R. 2002. Context modulates processing of speech sounds in the right auditory cortex of human subjects. *Neurosci Lett*. 331:91-94.
- Liebethal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. *Cereb Cortex*. 15:1621-1631.
- Logan JS, Lively SE, Pisoni DB. 1991. Training Japanese listeners to identify English /r/ and /l/: a first report. *J Acoust Soc Am*. 89:874-886.
- Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 54: 1001-1010.
- Mangasarian OL, Musicant DR. 2000. LSVM Software: active set support vector machine classification software. [Internet] Available at <http://www.cs.wisc.edu/dmi/lsvm/>. Accessed 5 April 2009.
- Mangasarian OL, Musicant DR. 2001. Lagrangian support vector machines. *J Mach Learn Res*. 1:161-177.
- McCandliss BD, Fiez JA, Protopapas A, Conway M, McClelland JL. 2002. Success and failure in teaching the /r/-/l/ contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn Affect Behav Neurosci*. 2:89-108.
- Mesgarani N, David SV, Fritz JB, Shamma SA. 2008. Phoneme representation and classification in primary auditory cortex. *J Acoust Soc Am*. 123:899-909.
- Miyawaki K, Strange W, Verbrugge R, Liberman AM, Jenkins JJ, Fujimura O. 1975. An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Percept Psychophys*. 18:331-340.
- Mourao-Miranda J, Friston KJ, Brammer M. 2007. Dynamic discrimination analysis: a spatial-temporal SVM. *Neuroimage*. 36:88-99.
- Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 15:1-25.
- Obleser J, Eisner F, Kotz SA. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci*. 28:8116-8123.
- Oxford Centre for Functional Magnetic Resonance Imaging of the Brain 2007. FSL. [Internet] Available at <http://www.fmrib.ox.ac.uk/fsl/downloads/>. Accessed 5 April 2009.
- Pessoa L, Padmala S. 2007. Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cereb Cortex*. 17:691-701.
- Poline JB, Worsley KJ, Evans AC, Friston KJ. 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage*. 5:83-96.
- Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. *Science*. 310:1963-1966.
- Pruitt JS, Jenkins JJ, Strange W. 2006. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *J Acoust Soc Am*. 119:1684-1696.
- Raizada RDS, Poldrack RA. 2007. Selective amplification of stimulus differences during categorical processing of speech. *Neuron*. 56:726-740.
- Remedios R, Logothetis NK, Kayser C. 2009. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J Neurosci*. 29:1034-1045.
- Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*. 123:2400-2406.
- Serences JT, Boynton GM. 2007. Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*. 55:301-312.
- Tricomi E, Delgado M, McCandliss B, McClelland J, Fiez J. 2006. Performance feedback drives caudate activation in a phonological learning task. *J Cogn Neurosci*. 18:1029-1043.
- Tsujimura N. 2007. An introduction to Japanese linguistics. 2nd ed. Malden (MA): Blackwell.
- Underbakke M, Polka L, Gottfried TL, Strange W. 1988. Trading relations in the perception of /r/-/l/ by Japanese learners of English. *J Acoust Soc Am*. 84:90-100.
- Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*. 281:1188-1191.
- Wang Y, Xue G, Chen C, Xue F, Dong Q. 2007. Neural bases of asymmetric language switching in second-language learners: an ER-fMRI study. *Neuroimage*. 35:862-870.
- Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci*. 10:685-686.
- Zhang Y, Kuhl PK, Imada T, Kotani M, Tohkura Y. 2005. Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage*. 26:703-720.