Cell
PRESS

Neuron
Article

# Selective Amplification of Stimulus Differences during Categorical Processing of Speech

**Rajeev D.S. Raizada**[1,*] **and Russell A. Poldrack**[2]
[1]Institute for Learning and Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195, USA
[2]UCLA Department of Psychology, Department of Psychiatry and Biobehavioral Sciences, and Brain Research Institute, Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA
*Correspondence: raizada@u.washington.edu
DOI 10.1016/j.neuron.2007.11.001

## SUMMARY

**The brain's perceptual stimuli are constantly changing: some of these changes are treated as invariances and are suppressed, whereas others are selectively amplified, giving emphasis to the distinctions that matter most. The starkest form of such amplification is categorical perception. In speech, for example, a continuum of phonetic stimuli gets carved into perceptually distinct categories. We used fMRI to measure the degree to which this process of selective amplification takes place. The most categorically processing area was the left supramarginal gyrus: stimuli from different phonetic categories, when presented together in a contrasting pair, were neurally amplified more than two-fold. Low-level auditory cortical areas, however, showed comparatively little amplification of changes that crossed category boundaries. Selective amplification serves to emphasize key stimulus differences, thereby shaping perceptual categories. The approach presented here provides a quantitative way to measure the degree to which such processing is taking place.**

## INTRODUCTION

Although functional neuroimaging studies have yielded a great deal of information about the human brain, the activation maps that are the end result face a fundamental problem. Such maps can show that a specific area is active during a given task, but they do not tell us what type of information processing is reflected by that activity. Neural activity does not come with labels attached.

In the present study, we sought to measure the degree to which a specific computational task is being carried out across the brain: the selective amplification of behaviorally significant stimulus differences. An ideally selective amplification would exaggerate only a small set of differences along a given stimulus dimension and would suppress all the others. This is exactly the case in categorical perception.

We drew our stimuli from speech perception, the domain in which categorical perception was first demonstrated (Liberman et al., 1957). The stimulus continuum consisted of speech sounds, with the phoneme /ba/ at one end and /da/ at the other. These two stop consonants differ only in place of articulation, with the articulatory difference reflected by different starting frequencies for the transitions of the second and third formants. A continuum of ten stimuli was synthesized, evenly spread along the /ba/-/da/ continuum. These stimuli will be referred to by their positions along the continuum, with "1" being the stimulus at the far /ba/ end and "10" being the extreme /da/.

Consider two stimuli along that continuum, e.g., items 4 (an impure mix of /ba/ and /da/, but closer to the /ba/ end) and 7 (an impure stimulus, closer to /da/). If these stimuli are presented back-to-back as a contrasting pair, 4-then-7 (henceforth written as 4&7), then we can measure the neural response elicited by the pair, using fMRI. Comparing the neural activity elicited by the contrasting pair to the sum of the activities elicited by its constituent subparts, there are three possibilities. The neural response to the pair may be equal to the sum of its parts, or, in the present shorthand, $4\&7 = 4 + 7$. This result would suggest that the two stimuli do not neurally interact. Alternatively, the response may be less than the sum of its parts, e.g., $4\&7 < 4 + 7$. One circumstance that would produce this result would be if a single population of neurons is responding to both stimuli, first to 4 and then, more weakly due to neural adaptation, to stimulus 7. In that situation, the neurons would be treating changes along the 4-to-7 section of the stimulus continuum as invariant (Grill-Spector and Malach, 2001). However, other explanations are also possible: the neurons may respond poorly to any sequence of stimuli, habituating nonspecifically, or the reduced fMRI response to the AB pair may be due to rate-limited blood flow rather than neural effects (Boynton and Finney, 2003).

In this study, we looked for cases in which the neural response is greater than the sum of its parts, e.g., $4\&7 > 4 + 7$. A supralinear response of this kind suggests that an active amplification process is occurring. However, such an amplification is of interest only if it is specific:

such increased activity must be elicited not by just any pairing of two stimuli along the continuum, but only by stimuli that lie on opposite sides of a particular boundary. To address this, we need to measure neural responses along the length of the continuum. If a brain area is found not only to amplify stimulus differences but also to do so with specificity, then a third question arises: is that neural amplification correlated with perception and behavior?

The phenomenon of categorical perception has all three properties: stimulus differences that do not cross the boundary are suppressed, whereas differences that straddle the category boundary are amplified, with a stimulus on one side of the boundary being perceived as very different from one on the other side, even though the two may be acoustically quite similar. For a continuum of auditory phonetic stimuli, a brain area that responds in this manner is processing the sounds as something more than just raw acoustic waveforms. This need not imply that the stimuli are being processed fully as speech, since nonhuman primates also show sensitivity to phonetic boundaries (Kuhl and Padden, 1983). However, selective amplification of this kind is likely to be a necessary property of any neural coding system that is rich enough to represent phonetic categories.

## Experimental Aims and Design

The aim of this study was to measure the degree to which brain activity parallels behavior, where the neural process is the selective amplification of stimulus differences and the behavioral phenomenon is categorical phonetic perception. Therefore, to find brain areas that process speech sounds categorically involves two steps. First, behaviorally measure how the subjects perceive stimulus differences along the continuum, to see which differences result in a change of phonetic category and which do not. Second, look for brain activity that selectively amplifies those stimulus differences in a parallel manner, such that differences that cause a phonetic change are neurally amplified, whereas nonphonetic differences are not.

Before laying out the details of the experimental design, it may be useful to outline which properties are intended to be measured by our behavioral and neural tests and which properties are not. In particular, the aim of the behavioral tests is to measure the degree to which a particular stimulus difference is perceived as a phonetic difference, e.g., how often a subject hears the stimulus pair 4&7 as consisting of two different phonemes. This same/different phoneme question is subtly different from asking what a subject's sensitivity is to a given acoustic difference. Although acoustic discriminability and phonetic difference tend to go hand-in-hand, a link that is at the core of categorical perception, the two are not identical; acoustic differences can be discriminable even without constituting phonetic differences, and phonetic differences do not always aid acoustic discrimination (Gerrits and Schouten, 2004). In the present study, we are seeking neural amplification that tracks phonetic differences. Although other stimulus differences may exist that are acoustically discriminable without being phonetically distinct, our fMRI measures do not attempt to track them.

The goal of the fMRI tests is to determine the degree to which the differences between stimuli are amplified, as we wish to find brain areas that specifically amplify differences that cross the perceptual category boundary. Measuring the response to the difference between two stimuli of course involves also measuring the responses to the two stimuli themselves, but it requires an extra step. For each contrasting stimulus pair, we must calculate two measures: the activation elicited by pairing the different stimuli together, and the activation elicited by the same individual phonemes recombined into noncontrasting identical pairs. For example, for the stimuli 4 and 7, we would measure the responses to 4&7 and 7&4 and the responses to 4&4 and 7&7. By subtracting the second measure from the first, i.e., (4&7 + 7&4) – (4&4 + 7&7), we are left with the amount of stimulus amplification caused by the contrast itself, over and above the activation elicited by the constituent subparts.

Although such a subtraction, of the form (A&B + B&A) – (A&A + B&B), provides us with a measure of how the pairwise difference between A and B is amplified, it does not tell us what the overall pattern of amplification is along the stimulus continuum. In particular, we are interested in finding patterns of selective amplification that would signify the occurrence of categorical processing: such processing would amplify not just any acoustic difference, but specifically only those differences that the subject perceives as phonetic. To search for such processing, we can compare each subject's behavioral same/different scores to their levels of neural amplification for each A/B pair. If the stimulus differences that a given brain area most strongly amplifies are the ones that the subject perceives as most different, then that constitutes evidence linking that brain area to perception.

More precisely, the best match between behavior and a given brain area will occur when the subject's behavioral same/different curve is parallel to that area's neural amplification curve. We can construct a general linear model contrast to measure this by weighting each neural amplification value, of the form (A&B + B&A) – (A&A + B&B), by the percentage of "different" responses that the subject made when presented with the A/B pair, and summing across all such three-step pairs.

The overall construction of this contrast is illustrated in Figure 3, and the arithmetical details are described in the Experimental Procedures section. A given voxel in the brain will score a large value on this contrast if it amplifies stimulus differences that cross the person's perceptual phonetic category boundary and also fails to amplify differences that do not cross the category boundary. The brain, at the location of that voxel, is processing the sounds categorically. In order to score a large value, the voxel therefore needs to meet two conditions simultaneously, one neural and one behavioral. The neural condition is that it must amplify some differences along the

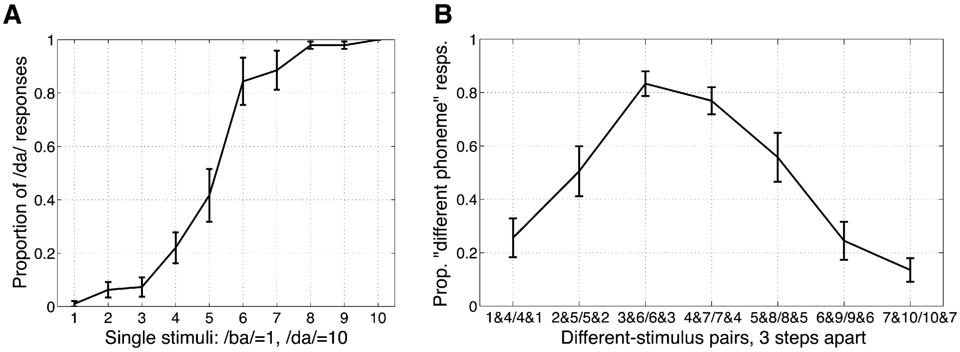

**Figure 1. Subjects' Psychophysical Responses to the /ba/-/da/ Continuum Stimuli, Tested Outside the Scanner**

(A) Subjects judged stimulus 1 to be /ba/ and 10 to be /da/ with very high consistency.

(B) Subjects showed the classical categorical perception effect of classifying stimulus pairs as different if they straddle the category boundary, and as the same if they both fall on the same side of the boundary. Error bars in these and in all other plots show the across-subjects standard error of the mean (n = 12).

phonetic continuum. The behavioral condition is that those differences that are neurally amplified must also be the ones that are perceptually amplified by crossing the phonetic category boundary.

### Evidence for Amplification versus Evidence for Adaptation

It is important to consider a difference between the logic of the present design and that of more typical adaptation-fMRI designs. In studies seeking evidence for the occurrence of neural adaptation, the experimentally meaningful signal change is a reduction in neural activity. The presence or absence of such a reduction serves as the logical starting point for claims about neural function (Grill-Spector and Malach, 2001).

In the present study, the reasoning follows a different path. The signal change that, we will argue, is demonstrated in the data below is not a reduction in neural activity, nor is it an adaptation-release-induced absence of such reduction. Instead, it is an active amplification of neural activity, in particular of activity elicited by stimulus pairs that straddle the phonetic category boundary. The essence of such amplification is that it yields strong overall activation out of weak raw ingredients. These raw ingredients are the individual stimuli that are combined together into the contrasting pair. The less activation these stimuli individually elicit, the weaker are the raw ingredients and the more neural amplification is required if the pair is to elicit strong activation when presented together.

The logic underlying neuroimaging of selective amplification could be viewed as the methodological converse of habituation-based methods that are used to look for invariances, such as fMRI-adaptation (Grill-Spector and Malach, 2001) and mismatch or oddball-based approaches (Celsis et al., 1999; Dehaene-Lambertz, 1997; May et al., 1999; Naatanen et al., 1997; Phillips et al., 2000; Zevin and McCandliss, 2005).

## RESULTS

### Behavioral Testing

In behavioral testing, subjects were presented with the synthesized speech stimuli and were asked to classify them as either /ba/ or /da/. As is shown in Figure 1A, subjects judged stimulus 1 to be /ba/ and 10 to be /da/ with very high consistency. The subjects' perceptual crossover point between /ba/ and /da/ was clustered around stimulus 5. In order to test for sensitivity to phonetic boundaries that is independent of raw acoustic differences, subjects were presented with stimulus pairs separated by three steps on the /ba/-/da/ continuum, e.g., 1&4, 6&3, 7&10, etc., and asked to say whether they sounded like the same phoneme or different phonemes. Randomly interspersed among the three-step contrasting pairs were equal numbers of pairs of identical stimuli (e.g., 1&1, 4&4, 6&6, etc.) to serve as controls. To prevent possible auditory masking effects, the two stimuli within a pair were separated by 500 ms of silence.

As is shown in Figure 1B, subjects' psychophysical responses showed the classical categorical perception effect of classifying stimulus pairs as different if they straddle the category boundary, e.g., 4&7 and 3&6, and as the same if they both fall on the same side of the boundary, even though the acoustic difference is constant across all pairs. The stimulus ordering was not an effect of interest for the purposes of this study, so responses to 4&7 and 7&4, etc., were collapsed together.

Note that the quantity plotted along the y axis is the proportion of times the subjects said that a given three-step pair sounded as though it consisted of two different phonemes, i.e., one /ba/ and one /da/. In signal-detection terms, this is the number of "hits." This quantity was used to construct a behaviorally weighted fMRI contrast, as described in detail below. Because the aim of the study was to see how stimulus differences that count as phonetic differences are neurally amplified, the proportion of

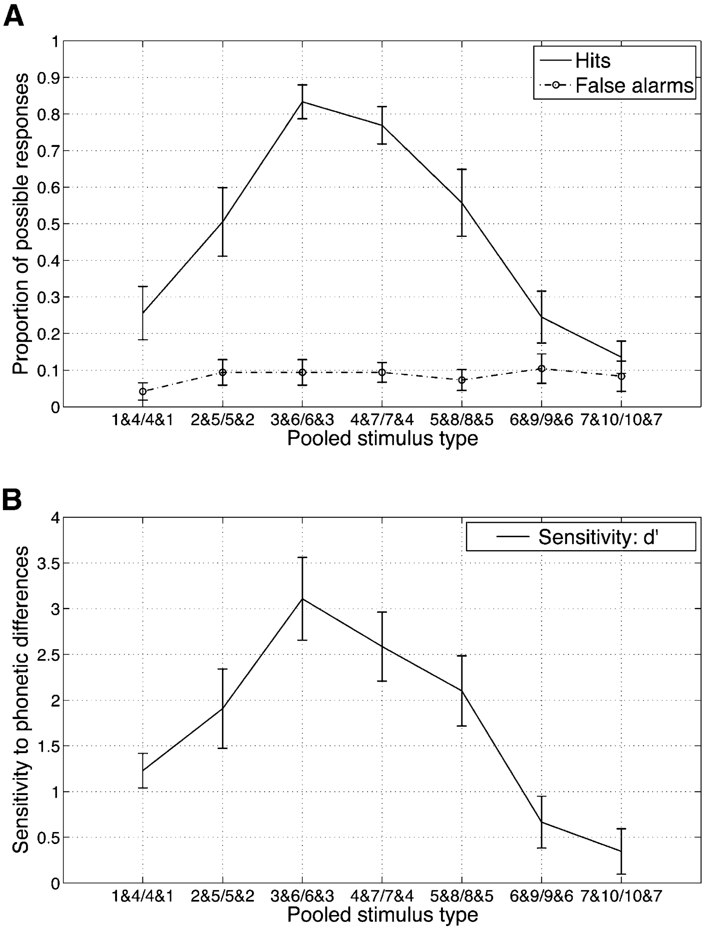

**Figure 2. Sensitivity, or d′, of the Subjects in the Same/Different Phoneme Identification Task**

The false-alarm rates remained relatively constant across stimulus conditions, indicating that the subjects did not show any systematic pattern of response bias. Error bars show the across-subjects standard error of the mean (n = 12).

"different phoneme" responses was the behavioral measure used to weight the fMRI data. Sensitivity, or d′, showed a similar pattern, consistent with the fact that false-alarm rates remained relatively constant across stimulus conditions (see Figure 2). However, d′ was not used to construct the fMRI contrast, as the resulting measure would be influenced by false-alarm neural processing of identical stimuli mistakenly perceived to be different. Although such false-alarm processing is of considerable interest, it was not the focus of the present study.

## fMRI Scanning

The same stimuli as were used in the behavioral testing were presented to the subjects in an event-related fMRI design in order to identify brain areas that selectively amplified the stimulus differences in a way that paralleled subjects' categorical perception. To prevent auditory masking by scanner noise, all stimuli were played during silent gaps between scan volumes, using the MRI technique of "clustered volume acquisition."

The neural process that we were interested in was categorical perception, as opposed to decision making, attention, motor-responses, and so on. Although such processes will to some unavoidable extent occur spontaneously, we sought to minimize them by not giving the subjects any phonetic discrimination task while they were in the scanner. Instead, and in order to maintain alertness in the subjects, we randomly interspersed some infrequent catch trials among the three-step pairs and identical-pairs. In these catch trials, both phoneme stimuli were the same, e.g., 6&6, but the second stimulus in the pair was quieter. The subjects' task was to press a button whenever such a trial occurred. This ensured that the subjects listened to all the stimuli, without needing to attend to any phonetic features. In order to construct the behaviorally weighted contrast illustrated in Figure 3, we of course needed to collect the subjects' behavioral identification and discrimination curves at some point. These behavioral tests, described above, were conducted outside of the scanner, a few minutes after each subject had finished their scan.

## Categorically Processing Regions

The behaviorally weighted contrast was constructed in order to quantify the degree to which any given brain area is processing speech categorically, such that phonetic differences are selectively amplified and nonphonetic differences are not. After having calculated the contrast for each of the 12 subjects, these contrast images were passed into a random-effects group analysis. Six regions

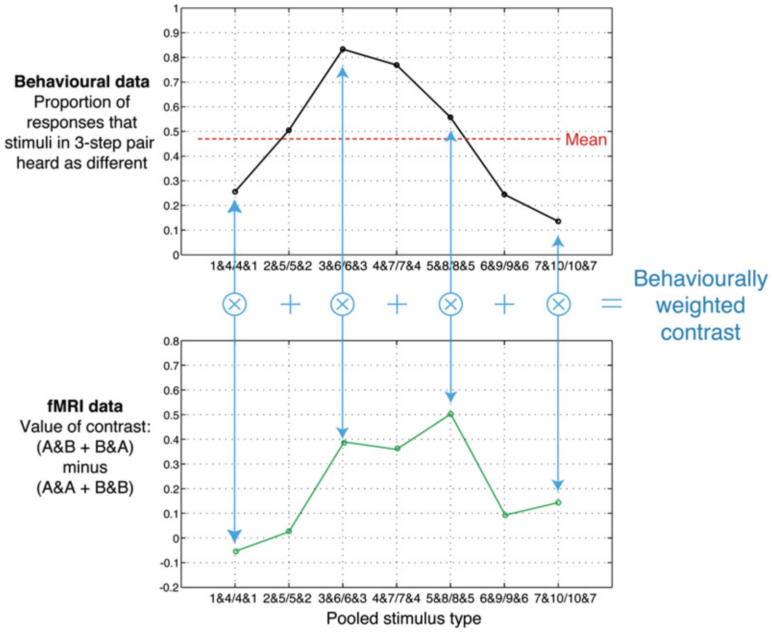

**Figure 3. Construction of the Behaviorally Weighted fMRI Contrast**

Schematic of how the behaviorally weighted fMRI contrast was constructed, to quantify how well a given voxel's pattern of neural amplification matches with the subject's behaviorally measured perception of phonetic differences. For visual clarity, this schematic diagram shows weighting arrows for only four of the points along the curve; for the construction of the contrasts, all seven points were used.

were found to process speech categorically, as measured by this selective-amplification contrast (corrected for multiple comparisons at the cluster level, at $p < 0.05$). The most significant of these activated regions was the left inferior supramarginal gyrus (shown in Figure 4A), consistent with lesion studies showing that the integrity of this region was more predictive than any other area of preserved ability to discriminate phonemes (Caplan et al., 1995). Although the supramarginal gyrus is less often thought of as involved in phonetic processing than the nearby Wernicke's area, a number previous functional imaging studies have highlighted this area as important (Benson et al., 2001; Callan et al., 2003; Jacquemot et al., 2003).

The other regions showing this phonetic selective-amplification effect are not ''classical speech areas,'' although some have been previously implicated in higher-order aspects of speech perception. Activation was found in the right cerebellum (Figure 4B), consistent with the results of Ackermann and colleagues (Mathiak et al., 2002), who found that a nearby right cerebellar locus was sensitive to differences in voice onset time, as opposed to the /ba/-/da/ contrast based on place of articulation. The anterior cingulate cortex also showed the phonetic-adaptation release effect (Figure 4C). This activity may correspond to attentional alerting to sounds crossing the perceptual category boundary (Benedict et al., 2002). We also found a large activation cluster in the left posterior parietal cortex (Figure 4D), extending down the intraparietal sulcus. The more inferior part of this activation cluster is close to a left parietal/supramarginal area reported by Celsis et al. (1999) to be sensitive to phonetic change. Other areas that we found to treat sounds as the same or different depending upon whether they crossed the perceptual category boundary were the left middle frontal cortex (Figure 4E) and the right prefrontal cortex (Figure 4F).

By looking for regions that neurally amplify stimulus differences crossing the category boundary, it is possible, we have argued, to identify brain areas that are carrying out categorical processing. A simpler question is whether these regions respond more to the three-step pairs that are perceived as different than to the three-step pairs that are perceived as the same, i.e., whether these regions respond more to across-category than to within-category pairs. Such a pattern of responses would be expected from categorically processing regions, although it would not in itself be sufficient evidence that categorical processing is taking place, as it does not distinguish between activation due to a stimulus pair's crossing of the category boundary and activation elicited individually by the separate phonemes that together make up the pair. Nonetheless, the comparison is one worth making, as a ''reality check'' that our six categorically processing regions are behaving as they would be expected to. The relevant data, verifying that the expected behavior does indeed take place, is shown in Figure S1 of the Supplemental Data available with this article online.

Having found these categorically processing regions, we wished to ask how their phonetic representations were structured, and to probe in more detail how they related to perception. In particular, we investigated how the continuum of individual stimulus types from /ba/ to /da/ was represented, the pattern of how stimulus changes along that continuum were amplified or suppressed, and the relation between this neural activity and people's perceptual discrimination.

## Neurometric Curves

To that end, we derived ''neurometric curves,'' plotting how neural activity varies as a function of position along the stimulus continuum, and compared them to subjects'

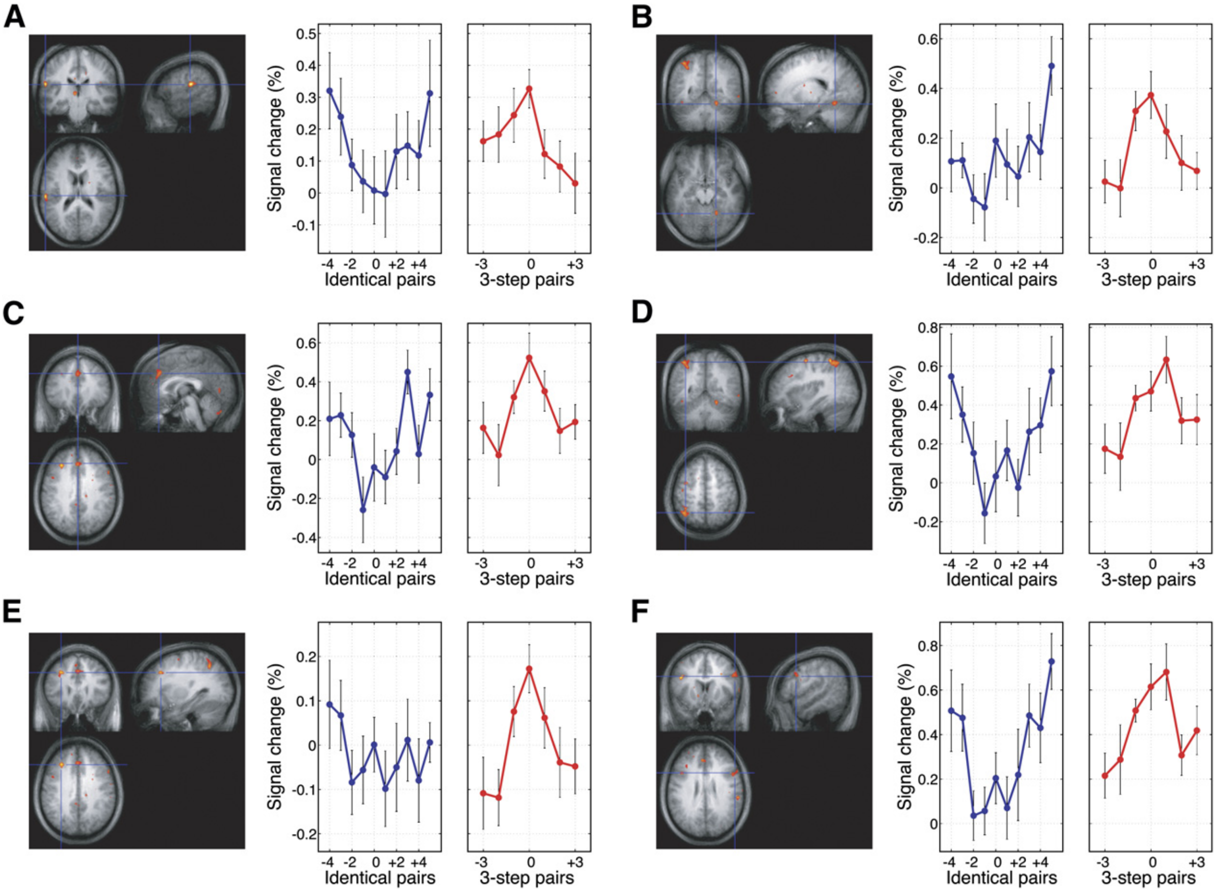

**Figure 4. The Six Brain Areas that Were Found to Process Speech Categorically**

The brain images show the results of the phonetic selective-amplification contrast, overlaid on the 12 subjects' mean anatomical scan. The corresponding neurometric curves are shown for each region. Plotted in blue is the neurometric curve of ''position responses'' to identical-stimulus pairs, aligned with respect to the perceptual /ba/-/da/ crossover point for each subject. Plotted in red are the ''contrast responses'' to three-step stimulus pairs, aligned with respect to the peak of each subject's behavioral same-different curve. (A) Left supramarginal gyrus, (B) right cerebellum, (C) anterior cingulate, (D) left parietal cortex, (E) left middle frontal cortex, (F) right prefrontal cortex. Error bars show the across-subjects standard error of the mean (n = 12).

psychometric curves, measured behaviorally. We derived two types of neurometric curves: ''position curves,'' plotting the responses to stimuli at individual locations along the continuum, and ''contrast-response'' curves, plotting the responses to pairs of stimuli that were three steps apart. The neurometric curves for the six categorically processing areas are shown in Figure 4. Note that these curves plot the actual BOLD responses, without any behavioral weighting.

A challenge in relating neurometric curves from multiple subjects to behavior is the fact that a given physical stimulus can be perceived differently by different people. Although, as Figure 1 shows, the /ba/-/da/ crossover points and same-different peaks showed fairly high consistency across subjects, one person's /ba/ nearer the to center of the continuum can be another person's /da/. Thus, in order to average like with like, each individual's neurometric curve was plotted as a function of stimulus position with respect to each subject's behavioral threshold, and it was these perceptually aligned curves that were averaged across subjects.

Thus, for the neurometric curve of ''position responses'' to the identical-stimulus pairs located at different positions along the /ba/-/da/ continuum (1&1, 2&2, etc.), each subject's neurometric curve was aligned with respect to the position on the continuum at which that individual behaviorally crossed over from perceiving /ba/ to perceiving /da/. Therefore, the ''0'' point on the x axis for these curves is the phonetic category boundary for every individual subject. The negative x axis values to the left are the /ba/ sounds, becoming progressively perceptually purer as they move away from the category boundary, and the positive x axis values to the right are the /da/ sounds. Thus, a V-shaped curve indicates that an area responds weakly to stimuli near the category boundary and strongly to phonetically purer stimuli at the category centers.

For the neurometric curve of ‘‘contrast responses’’ to three-step pairs (1&4, 9&6, etc.), the curves were aligned with respect to the peak of each subject’s behavioral same-different curve, so that the ‘‘0’’ point again corresponds to the phonetic category boundary. Here, an inverted-V shape indicates that an area responds strongly to stimulus pairs that straddle the category boundary, and which are hence perceived as different, and that the area responds weakly to stimuli that are perceived as the same.

A key aspect of the categorically processing areas’ neurometric curves, as seen in Figure 4, is that they respond very weakly to individual stimulus types near the category boundary, but nonetheless respond strongly to contrasting pairs made out of those same stimuli. For example, in Figure 4A, the left supramarginal gyrus averages a response close to zero for pairs of identical stimuli near to each subject’s category boundary (the x axis zero in the blue curve of ‘‘position-responses’’ to identical pairs). However, the three-step pairs that straddle the category boundary are constructed out of exactly these stimuli drawn from near the middle of the continuum, yet they elicit the strongest responses (as shown by the sharp peak at the category boundary in the red ‘‘contrast-response’’ curve). The neurometric curves thus graphically illustrate the fact that large neural responses to behaviorally important stimulus differences are amplified out of much smaller responses to their constituent subparts.

The converse pattern also holds true: many of the areas, again most notably the left supramarginal gyrus, respond strongly to individual stimulus types at the ends of the /ba/-/da/ continuum. However, contrasting three-step pairs made out of these phonemes induce only weak neural responses (the low activity at the far left and far right ends of the red ‘‘contrast-response’’ curves), even though the pairs are made out of stimuli that induce strong responses individually. The selectivity of the brain’s amplification here shows its flip side: just as contrasts that straddle the category boundary are selectively amplified, contrasts that do not straddle the boundary are selectively suppressed.

## Neural Amplification and the Shape of Neurometric Curves

Neural amplification of a stimulus difference, as operationalized here, occurs when a brain area responds strongly to a contrasting three-step pair, and yet at the same time responds weakly to the individual stimuli out of which that contrasting pair is built. In terms of the neurometric curves, this corresponds to having a peak at the category boundary for the red ‘‘contrast-response’’ curves and a trough at the boundary for the blue ‘‘position-response’’ curve. Such amplification is maximally selective if stimulus differences that do not cross the category boundary are suppressed. This corresponds to having low values at the outer ends of the red ‘‘contrast-response’’ curves and high values at the ends of the blue ‘‘position-response’’ curves. Thus, a region that has an inverted-V-shaped ‘‘contrast-response’’ curve and an upright-V ‘‘position-response’’ curve will score well in the selective-amplification contrast. The left supramarginal gyrus, whose curves are plotted in Figure 4A, provides an example of such behavior. However, such a clean pair of upright and inverted Vs is not the only way to score well on the contrast. For example, the left middle frontal cortex, shown in Figure 4E, produced significant categorical processing activation, without showing a marked upright-V ‘‘position-response’’ pattern. It made up for this with the strength of its inverted-V ‘‘contrast-response’’ curve. Thus, there are many different ways of scoring high on the categorical-processing contrast, although there is only one way of scoring perfectly. As the diversity of different neurometric curve shapes shown in Figure 4 illustrates, not all categorically processing regions are the same.

Because the neurometric curves plot the actual BOLD responses, without any behavioral weighting, the V-shapes and other patterns described above are totally independent of the subjects’ behavioral responses. The positions of the subjects’ perceptual category boundaries were used only to align their individual curves with each other along the x axis and did not affect the strength of neural activity plotted along the y axis.

## Neural Amplification and Comparing the Whole to the Sum of Its Parts

In Figure 4, the neurometric curves for the three-step pairs are plotted separately (the red curves) from those for the same constituent stimuli rearranged into identical pairs (the blue curves). However, in order more directly to visualize the difference in activation between the contrasting and noncontrasting stimulus pairings, it is also useful to overlay the curves on top of each other. These overlaid curves for the categorically processing regions are shown in Figure 5. In these overlaid plots, the red curves of responses to three-step pairs are the same as in Figure 4, but the blue curves now show the average neural responses to the identical-pairs made out of the corresponding constituent subparts. For example, where a red three-step curve shows the average of the responses to 4&7 and 7&4, the blue curve at the same x-position shows the average of the responses to 4&4 and 7&7.

The overlaid plots allow the occurrence of neural amplification to be directly visualized. Neural amplification takes place when the response to the contrasting pairs is greater than the response to those pairs’ constituent subparts, in other words, when the red curve is higher than the blue. As can be seen from Figure 5, such amplification takes place near the category boundary for all six of the categorically processing regions. Moreover, this amplification is selective, as it occurs only near the category boundary. Indeed, in some of the regions, the curves cross over, with the blue curve becoming higher than the red at positions near the edges of the stimulus continuum, suggesting that stimulus differences far away from the category boundary may be suppressed.

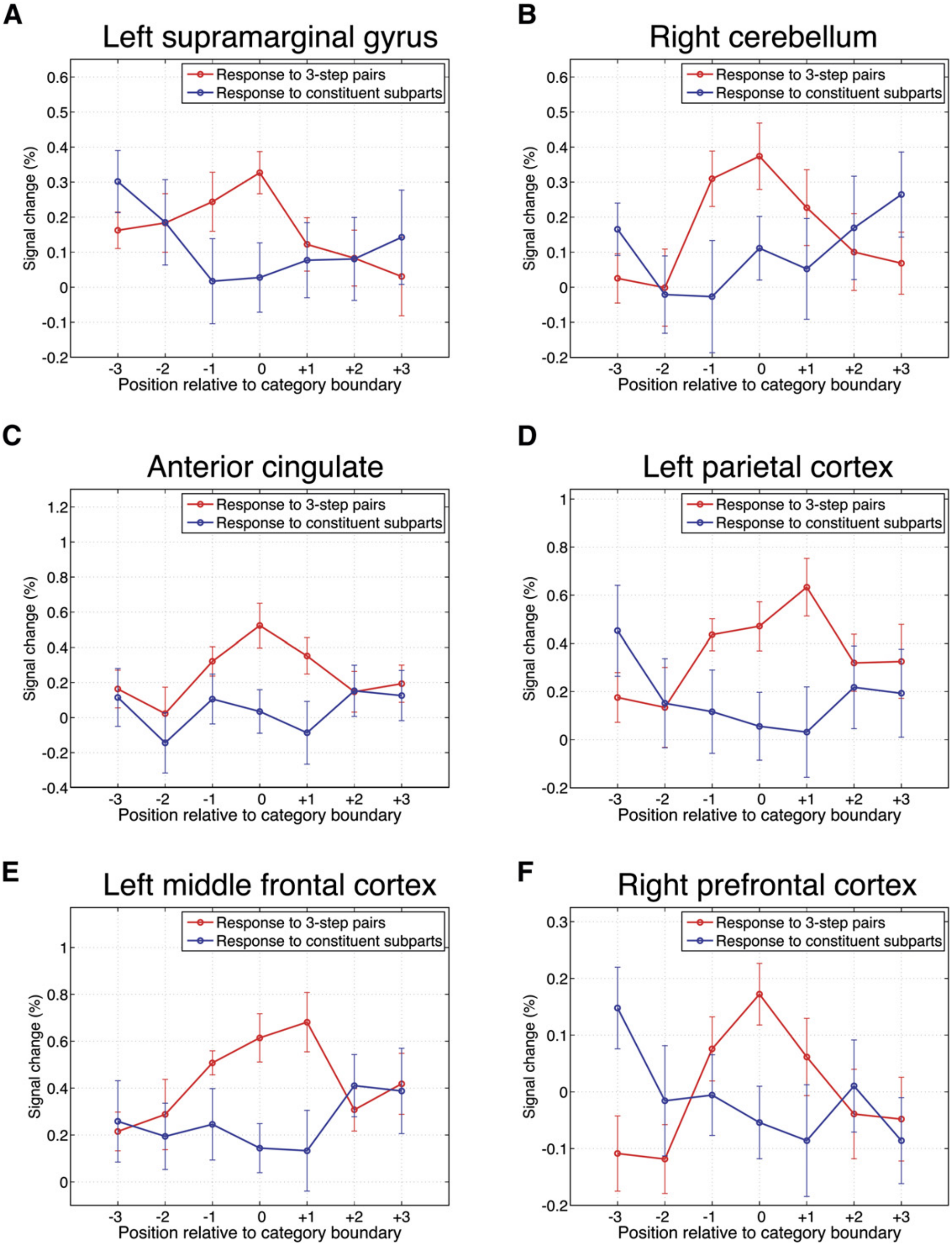

**Figure 5. Overlaid Neurometric Curves for Identical Pairs and Three-Step Pairs**

Neurometric curves for the six categorically processing regions, showing the same data as Figure 4, but now with the curves overlaid on top of each other, in order more directly to visualize the difference in activation between the contrasting and noncontrasting stimulus pairings. In these overlaid plots, the red curves of responses to three-step pairs are the same as in Figure 4, but the blue curves now show the average neural responses to the identical pairs made out of the corresponding constituent subparts. Error bars show the across-subjects standard error of the mean (n = 12).

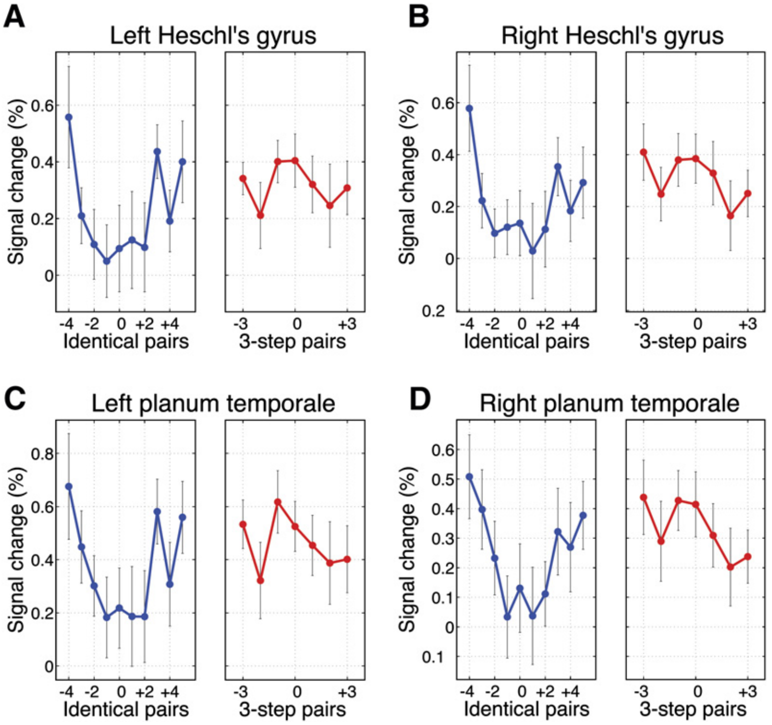

**Figure 6. "Position-Response" and "Contrast-Response" Neurometric Curves for the Lower-Level Anatomically Defined Regions of Interest**

These areas tend to be comparatively insensitive to the phonetic category boundary, as can be seen by the lack of a strong inverted-V pattern in the red "contrast response" curves. (A) Left Heschl's gyrus, (B) right Heschl's gyrus, (C) left planum temporale, (D) right planum temporale. Error bars show the across-subjects standard error of the mean (n = 12).

### Neurometric Curves in Lower-Level Auditory Cortical Areas

We were interested in exploring the neural representations not only of these categorically processing areas, but also in lower-level auditory areas in the superior temporal lobe: Heschl's gyrus, and the planum temporale (or PT). These areas are robustly activated by speech stimuli, but they failed to show the categorical-processing effect above, raising the question of which aspect of the speech signal they do encode. To address this, anatomical regions of interest (ROIs) of these areas were traced out on each subject's high-resolution T1-weighted anatomical images, on both the left and the right. The neurometric curves for these anatomically defined ROIs are shown in Figure 6.

The most notable aspect of the lower-level areas' neurometric curves is how much flatter their "contrast-response" curves are. The response to three-step stimulus pairs that straddle the category boundary is barely any greater than the response to contrasting pairs at the extreme ends of the /ba/-/da/ continuum. Although there is a hint of a small increase in activity at the category boundary, it is much weaker than in the six categorically processing areas shown in Figure 4. The overlaid neurometric curves for these lower-level areas, allowing a more direct visualization of the presence or absence of neural amplification, are shown in Figure S2. As can be seen from that figure, the responses to the three-step pairs are slightly stronger than the responses to their constituent subparts, but this effect is much less pronounced in these lower-level auditory areas than it is in the categorically processing areas such as the left supramarginal gyrus.

Although the neurometric curves provide, together with their error bars, a direct visual indication of these effects, they are not in themselves enough to show their statistical consistency across subjects. We therefore constructed *t*-contrasts, treating subjects as a random effect, and assessed them by applying Bonferroni correction to take into account the fact that ten ROIs were being examined. Of the six areas that processed speech categorically, shown in Figure 4, five of them showed a significant pattern of responding to the category boundary ($p < 0.05$, Bonferroni corrected). The only area that did not show this effect was the left parietal cortex. This confirms the prima facie observation from Figure 4 that the red neurometric curves show an inverted-V pattern, peaking at the category boundary, with this V-shape being less marked for left parietal (Figure 4D). In contrast, Heschl's gyrus and the planum temporale did not show significant responses to the category boundary, either on the left or the right ($p > 0.28$ for all four areas). This shows that these lower-level auditory areas are relatively insensitive to the phonetic category boundary. This statistically quantifies the visual observation from Figure 6 that the red three-step neurometric curves for these anatomical areas are relatively flat, failing to show a clear inverted-V pattern.

The overlaid neurometric curves for the lower-level auditory regions are shown in Figure S2, plotting the difference in activation between the contrasting and non-contrasting stimulus pairings. These plots confirm that the lower-level regions show much weaker levels of amplification, if any at all. Nonetheless, these areas do respond more to contrasting stimuli than to repeated identical stimuli. They do, therefore, show a preference for stimulus

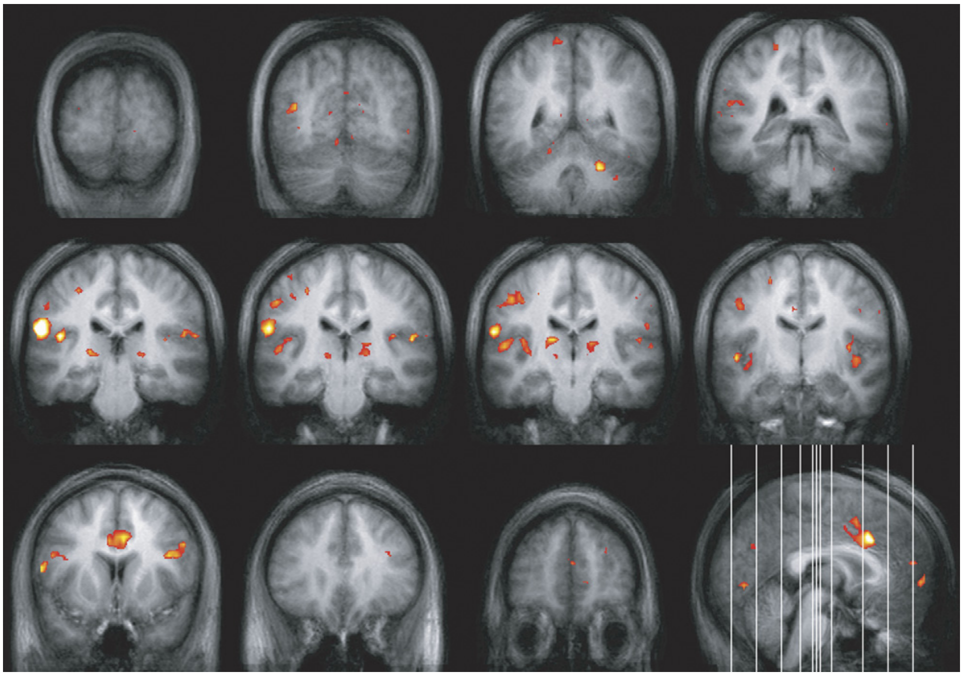

**Figure 7. Areas Whose Activation Time Courses Are Correlated with that of the Left Supramarginal Gyrus**
Results are from a random-effects analysis, thresholded at $p < 0.05$, corrected for multiple comparisons at the voxel level. The left supramarginal gyrus activation correlates with several areas, most notably left Heschl's gyrus, left planum temporale, bilateral medial geniculate nucleus, and the anterior cingulate.

change—a very general characteristic of the nervous system. Unlike the categorically processing regions, these lower-level areas exhibit a preference for stimulus change that is not so specifically focused on the phonetic category boundary.

A speech-related area that might have been expected a priori to process speech categorically, but which did not show this effect in our study, was Broca's area. A possible explanation for this is that our study was of the perception of isolated phonemes, whereas Broca's area may be involved more in the decomposition and combination of multiple phonemes (Gelfand and Bookheimer, 2003), as would occur when listening to or producing a continuous stream of speech. Similarly, it might have been expected that the planum temporale would have shown a stronger contrast response at the category boundary. However, the absence of such an effect need not imply that a brain area is insensitive to such boundaries. It could be that some small group of highly-tuned PT neurons responds vigorously to stimulus-pairs that straddle the category boundary, but that this localized firing induces too small a haemodynamic effect to be detected by fMRI. Another possibility is that the PT may represent category boundaries in its patterns of neural communication with other brain areas, rather than in purely local activity. Although fMRI may be able to detect some varieties of neural mechanism, there are doubtless many more that remain beyond its grasp.

### Functional Connectivity between Lower- and Higher-Level Areas

This finding raises the question of what the functional relations might be between these lower- and higher-level areas. In order to address this, we carried out a functional connectivity analysis (see the Experimental Procedures section), looking across the whole brain for voxels whose activation time courses correlated with that of the left supramarginal gyrus ROI, which was the region showing the strongest tendency to process speech sounds categorically. As can be seen from the results of this analysis in Figure 7, the supramarginal region correlates strongly with several brain areas, most notably left Heschl's gyrus and the left planum temporale, with some correlation also present in the corresponding areas on the right. This suggests that although these lower-level superior temporal regions are not in themselves very sensitive to the phonetic category boundary, they are in close communication with the region that is maximally sensitive to this boundary: the left supramarginal gyrus. Auditory areas at an even lower level also correlated with the supramarginal gyrus, namely the bilateral medial geniculate nucleus. The supramarginal activation time course also correlated with some of the other categorically processing regions, in particular the anterior cingulate and the right cerebellum, suggesting that these regions may constitute a functional network for the categorical processing of speech.

### Under What Conditions Is Amplification Demonstrated?

The essence of neural amplification is that its yields strong overall activation out of weak raw ingredients. In order to demonstrate that such amplification is taking place, we must isolate the amount of stimulus amplification caused by the contrast between two stimuli, over and above the activation elicited by the stimuli themselves. As was described above, we addressed this by comparing the activation elicited by pairs of contrasting stimuli, e.g., 1&4 and 4&1, against the activation elicited by the same individual phonemes recombined into noncontrasting identical pairs, in this instance 1&1 and 4&4.

#### *Using Paired Stimuli versus Using Isolated Stimuli*

A possible alternative method would have been to compare the three-step pairs against the activity elicited by isolated individual phonemes, e.g., 1 presented on its own, and 4 on its own. However, presenting isolated stimuli would have caused several potential psychological confounds. To avoid expectation-based effects, the paired and isolated stimuli would have had to be randomly intermingled, but then the frequent switches between paired and isolated stimulus presentations would have induced large attentional and alerting effects, which could have swamped the more subtle phonetic signals. A key aspect of categorical perception is that a pair of physically distinct stimuli, such as 1&4, may be perceptually very similar to a pair of physically identical stimuli, such as 4&4, since for most subjects both 1 and 4 lie on the same side of the /ba/-/da/ category boundary and hence would be perceived as sounding the same. By always using paired stimuli, we were able to access this type of effect.

#### *Inferring the Effects of Isolated Stimuli from Measurements of Paired Stimuli: Principles and Caveats*

Although the above considerations ultimately favor using paired rather than isolated stimuli, our use of paired stimuli raises a number of interpretive challenges.

In particular, the key test for whether neural amplification is occurring is whether the response to a boundary-crossing pair of stimuli is greater than the sum of the responses to its constituent subparts, for example whether the response to 4&7 is greater than the responses to 4 and 7 presented individually. However, for the reasons described above, all the stimuli were presented in pairs rather than individually. Thus, we are faced with the challenge of estimating the neural responses to isolated stimuli, which were not measured, from the responses to stimuli presented in pairs, which were measured.

This use of identical-pair stimuli raises a potential problem for the method: if a brain area habituates strongly to repeated identical stimuli, then the response to, say, 4&4, would be less than the response that would have been induced by presenting two isolated 4 stimuli. Although the 500 ms silence that we placed between the two stimuli would be more than sufficient to prevent auditory interference such as backward or forward masking (Gelfand, 2004; Smiarowski and Carhart, 1975; Wilson and Carhart, 1971), neural habituation can operate on a longer timescale and could still occur (Grill-Spector and Malach, 2001). If such adaptation did occur for the identical pairs, and did not occur for the three-step pairs, then an artifactual appearance of active amplification could be created, when the underlying process was in fact the absence of neural habitation.

There are two possible scenarios that could potentially undermine the claim that amplification takes place, which we will now consider. In one possibility, some kind of active inhibitory process suppresses the second member of an identical stimulus pair to the extent that its activation becomes negative. Under that hypothesis, our data permit two closely related but not identical interpretations: either the across-category stimulus pairs are actively amplified, or the within-category pairs are actively suppressed. In the second possibility, neural habitation occurs for the identical pairs to the extent that the second stimulus in the pair may be reduced to zero, but without any active inhibitory process pushing the activation into becoming negative. As we will argue below, even in this extreme habituative case our data still provide evidence that a genuine amplification takes place, i.e., that the response to a stimulus pair is indeed greater than the sum of the responses to its parts.

#### *Possibility 1: Identical Pairs Are Actively Suppressed to Become Negative*

Neural data from human MEG studies (McEvoy et al., 1997) and animal neurophysiology (Brosch and Schreiner, 1997) show that when stimuli are presented together in pairs the neural response to the second stimulus in the pair gets attenuated, but is not suppressed into being negative. This held true even when the ISI was shorter than the 500 ms used in the present study (for examples of the data from those studies, see Figure S7). Nonetheless, it is instructive to consider what our data would mean in the hypothesized scenario in which the second stimulus were suppressed far enough to become negative.

Under this hypothesis, our actual observation that stimulus pairs such as 4&7 and 7&4 elicit larger responses than 4&4 and 7&7 would not necessarily entail that the across-category three-step pairs had been amplified to become greater than the sum of their parts. This is because the observed small activation from an identical pair such as 4&4 might not have arisen from two small neural responses to each of the individual 4 stimuli. By hypothesis, the small response to the 4&4 pair could instead have been due to the first 4 in the pair producing a strongly positive response and then the second 4 in the pair producing a strongly negative response. This negative response would act to cancel out the large first response, thereby making the measured joint response be small overall.

On this hypothesis then, the raw ingredients of 4 and 7 in isolation could have been strong, even though the actually observed responses to 4&4 and 7&7 were weak. Thus, the observed strong response to 4&7 would not count as evidence for the amplification of a strong response out of weak raw ingredients.

Given the MEG and neurophysiological studies cited above, we consider this scenario to be empirically unlikely. It is, however, a logical possibility. In order to definitively rule out this possibility, it would be necessary to have independent and separate measures of the responses to each of the individual stimuli within a pair. The time resolution of fMRI does not permit this. The MEG and single-unit neurophysiology studies did carry out such measurements, and they suggest that this hypothesized positive-then-negative pattern of paired activation does not occur. However, the tasks and stimuli in those experiments were not completely identical to those in the present study, so this hypothesis remains logically open.

Under this hypothesis, then, identical pairs such as 4&4 would be actively suppressed. Such pairs are of course within-category, in virtue of being identical. Thus, in this scenario, there are two closely related possible interpretations of our data: either the across-category pairs are being actively amplified, or the within-category pairs are being actively suppressed. Indeed, both processes may be happening together. In either case, an active neural process is serving to sharpen the phonetic category boundary.

### Possibility 2: Identical Pairs Strongly Habituate, but Are Not Actively Suppressed

We now consider the empirically more plausible scenario in which the stimuli within the identical pairs strongly habituate, but without any active suppressive process pushing the stimuli into being inhibitory. In this circumstance, the "worst-case scenario," which would maximally tend to make actually strong isolated stimuli falsely appear to be weak when measured in pairs, would be the following: the second stimulus in an identical pair would experience 100% habituation. In that scenario, it would no longer be the case that 4&4 + 7&7 could be considered equal to the responses to its isolated subparts 4 + 4 + 7 + 7. Instead, under the hypothesized worst case 100% habituation, the responses to the second members of each of these identical pairs would be zero. Thus, in any circumstance up to and including this worst case, the response to 4&4 together would be greater than or equal to the response to the first 4 presented in isolation. In shorthand: $4 \leq$ 4&4. Therefore, $4 + 4 + 7 + 7 \leq$ 4&4 + 4&4 + 7&7 + 7&7 = 2 × (4&4 + 7&7). This means that even the case of 100% habituation restricted exclusively to identical-pair stimuli would produce the appearance of amplification by at most a factor of two.

Thus, a strict test of whether stimulus differences are neurally amplified is as follows: the contrasting three-step pairs must not only elicit more activity than the noncontrasting identical pairs made from the same raw ingredients, they must elicit at least twice as much activity.

In order to quantify this, let $X$ be equal to a voxel's average activation at a given position along the continuum of three-step pairs, e.g., the average activation elicited by the stimuli 4&7 and 7&4. Let $Y$ be equal to the voxel's average activation elicited by the corresponding identical pair stimuli, in this case 4&4 and 7&7. Then the strict test of whether neural amplification is taking place is to see whether $X > 2Y$, or equivalently $X - 2Y > 0$.

We calculated the value $X - 2Y$ for each of the categorically processing regions described above and also for the lower-level anatomically defined ROIs. The results for the categorically processing regions are shown in Figure S3, and those for the anatomical ROIs in Figure S4. In the categorically processing regions, it can be seen that near the category boundary the inequality is satisfied; stimulus differences that straddle the category boundary are indeed amplified.

For a much more detailed case-by-case breakdown of the above argument, interested readers may refer to Section 1 of the Supplemental Data.

### Possible Edge Effects in Adaptation Space and Evidence for Amplification

A related possible concern that might appear to undermine our amplification claim is the possible occurrence of edge effects in adaptation along the stimulus continuum.

By an "edge effect," we mean the following: in the present study, the stimuli vary along a continuum in which the manipulated dimension is the formant transition's starting frequency. Other studies of categorical perception have varied stimuli along different types of continua, notably studies of faces, which have manipulated stimuli along the dimension of facial identity (Jiang et al., 2006; Leopold et al., 2001; Rotshtein et al., 2005). A potential difficulty for all studies using such continua is that the stimuli differ not only in their positions along the continuum, but also in how far they are from the continuum's outer edges. This is a problem, because neurons can undergo adaptation not only by being presented with their maximally "preferred stimulus," but also by similar stimuli that occupy neighboring positions along the continuum. Such adaptation can even be induced by stimuli that are quite far away along the continuum; for example, Leopold et al. (2001) were able to bias people's perception of facial identity by presenting "anti-faces" on the opposite side of face-space.

Because of such nonlocal adaptation induction, "edge effects" may arise: stimuli at the center of the stimulus space will always tend to be exposed to more neural adaptation than stimuli at the edges. This is a simple consequence of the fact that a stimulus at the edge of the space can experience adaptation-inducing influences from only one direction, whereas stimuli near the center can receive such influences from all directions.

In the present study, the stimuli near the center of our phonetic continuum participate in more three-step pairs than do stimuli at the edge, with the result that these center stimuli are therefore presented more frequently. For example, stimulus 4 is a member of the pairs 1&4 and 4&7, whereas stimulus 1 is a member only of the pair 1&4. This raises the following key question: if such an edge effect were to lead to continuum-center stimuli undergoing additional adaptation, could this thereby undermine our claim to have demonstrated the occurrence of neural amplification?

This question can be answered in the negative. Indeed, to the contrary, if such additional adaptation were to occur, it would in fact strengthen the evidence that neural amplification is taking place. Recall that the hallmark of amplification is that it yields strong overall activation out of weak raw ingredients. If edge effects were to cause individual stimuli near the continuum center such as 4 and 7 to become neurally weakened by additional adaptation, then these raw ingredients would be even weaker than otherwise. If this were the case, then it would require even more amplification to account for our observed data, as these weak ingredients yield strong activation when presented together as 4&7 in a boundary-crossing three-step pair. This possible additional weakening of the continuum center stimuli requires that even greater amplification than otherwise must be taking place.

For further discussion of these issues, along with examples from our data illustrating the points, please see Section 2 in the Supplemental Data, and Figures S5 and S6.

## DISCUSSION

In order to study the structure of the brain's representations, it is necessary not only to record activation, but also to try to determine what type of information processing is reflected by the observed activation. Here, we used fMRI to measure the degree to which perceptually important stimulus changes were being specifically amplified, across the whole brain. In the present study, the occurrence of such specific amplification indicated that the sounds were being processed as something more than just raw acoustic waveforms, thereby providing the building blocks of a neural coding system that is rich enough to represent phonetic categories.

The finding that the left supramarginal gyrus is the most categorically processing region is consistent with other studies of that region (Benson et al., 2001; Callan et al., 2003; Caplan et al., 1995; Jacquemot et al., 2003) and also with recent studies revealing the broad and interconnected network of areas involved in speech perception (Blumstein et al., 2005; Griffiths et al., 2004; Hickok and Poeppel, 2000; Liebenthal et al., 2005). It is likely that different types of speech tasks may differentially recruit distinct regions in the brain's network of language processing areas. For example, Binder et al. (2004) found that a phoneme-identification task activated an area slightly anterior to Heschl's gyrus. The categorical processing probed in the present study is more closely related to discrimination than to identification tasks, possibly accounting for the different pattern of neural activity observed.

### Neural Amplification or Just BOLD Signal Amplification?

As with all fMRI studies, the BOLD signal measured here is only an indirect measure of the underlying neural activity. Thus, any inference made about neural amplification should be understood as depending upon the premise that the BOLD signal and neural activity are indeed correlated. Simultaneous neurophysiological and fMRI recordings in monkeys suggest that this premise is indeed valid (Logothetis et al., 2001), but nonetheless the fact that our measurements are of BOLD rather than direct neural activity should be borne in mind.

### Putting the "Selective" in Selective Amplification

A key aspect of the method proposed here, and one that is especially relevant to categorical perception, is that we were seeking not just neural amplification per se, but in particular selective amplification. Thus, an area such as the left supramarginal gyrus not only amplified differences between the phonetic stimuli, but moreover it specifically amplified only the differences that corresponded to crossing each subject's perceptual category boundary. This selectivity can be seen clearly in, for example, Figure 4A. Identical stimulus pairs that are near to the perceptual category boundary produce activation that is, on average, barely above zero. However, these same stimuli combined into three-step pairs induce large activation. Yet away from the category boundary, this pattern is totally reversed. This complete change in behavior caused by moving away from the perceptual boundary is the mark of categorical processing, and this is what puts the ''selective'' in selective amplification.

Perhaps the purest form of selective amplification is that which occurs in categorical perception: differences that cross a category boundary are amplified into maximum salience, and other differences that fail to cross the boundary are strongly suppressed. Many stimulus domains have been found to be processed categorically, including object-shape (Freedman et al., 2001), facial expressions (Etcoff and Magee, 1992), and color (Bornstein et al., 1976). Recent fMRI studies have debated whether face perception rests upon categorical processing along the dimension of identity (Rotshtein et al., 2005) or, alternatively, shape-tuned processing that lacks any explicit representation of category boundaries (Jiang et al., 2006). It is possible that studying the patterns of selective neural amplification involved in such processing, using the methods presented here, may be able to cast light on these issues.

The way in which the brain selectively amplifies stimulus differences can help to reveal how its representations of the world are structured. Such amplification can be said to be involved in a neural representation, as opposed to being just incidental activity, only if it is related to perception and behavior, a judgment that can be made by comparing neurometric curves to their psychometric counterparts. Used together, these tools can help to reveal when the brain sees the world in shades of gray, and when it sees in black-and-white.

## EXPERIMENTAL PROCEDURES

### Subjects

There were 12 subjects in all: 7 female, 5 male, age range 21–36. All were right-handed native English speakers and gave informed consent according to procedures approved by the Human Research Committee of Massachusetts General Hospital.

### Phoneme Stimuli

The individual /ba/ and /da/ stimuli were each 300 ms in duration with 500 ms of silence separating the two members of each stimulus pair. The stimuli were made using a SenSyn Klatt Synthesizer (Sensimetrics, Inc.) and varied in the second and third formants. The formant transitions lasted 150 ms and had the following start frequencies. /ba/: F2 = 1400 Hz, F3 = 2204 Hz; /da/: F2 = 2027 Hz, F3 = 2900 Hz; End: F2 = 1660 Hz, F3 = 2490 Hz. The fundamental frequency, F0, decreased linearly over time, from 144 to 108 Hz. The F1 formant went from 300–600 Hz in 0–50 ms, then to 690 Hz by 150 ms.

### Behavioral Testing

The subjects' psychometric curves were behaviorally measured outside the scanner a few minutes after the end of the scanning session. Testing was carried out in a quiet room, with the stimuli presented via headphones from a Macintosh laptop, with software written using the Matlab Psychophysics Toolbox (Brainard, 1997). Subjects were first given a stimulus-identification task in which they were presented with individual stimuli along the continuum and had to identify each one as either /ba/ or /da/. Each of the ten stimuli along the continuum was presented eight times, with the order randomly intermixed. Subjects had up to 2 s to respond to each trial.

After a brief break, the subjects were then given a same/different task in which they were presented with stimulus pairs separated by three steps on the /ba/-/da/ continuum, e.g., 1&4, 6&3, 7&10, etc., and asked to say whether they sounded like the same phoneme or different phonemes. Randomly interspersed among the three-step contrasting pairs were equal numbers of pairs of identical stimuli (e.g., 1&1, 4&4, 6&6, etc.) to serve as controls. To prevent possible auditory masking effects, the two stimuli within a pair were separated by 500 ms of silence. The ordering of stimuli within pairings was counterbalanced. Each of the three-step pairs was presented eight times, as were each of the identical pairs. The data analysis collapsed across within-pair ordering, pooling 3&6 and 6&3 together, etc., so that there were 16 instances of each contrast.

### fMRI Scanning

The fMRI was carried out in a Siemens 3T Trio scanner at the MGH-NMR Center, using a birdcage headcoil, a standard EPI BOLD pulse sequence, and a clustered volume acquisition with the following parameters: TR = 4 s, TA = 1.8 s, silent gap = 2.2 s, 500 ms interval between stimuli, and scanner-noise onset/offset 25 slices, 3.1 × 3.1 mm within plane resolution, 5 mm thick slices with a 0.5 mm skip, and descending slice-ordering. Each stimulus pair was presented in the middle of the 2.2 s clustered volume acquisition silent gap. In the scanner, sounds were played via nonmagnetic Koss electrostatic headphones, adapted for fMRI by Giorgio Bonmassar and Patrick Purdon.

The fMRI scans were subdivided into seven runs, with 104 volume acquisitions per run. 700 stimuli were present in all, using 100 stimuli per run. Of the 700 stimuli, 480 were phoneme trials, 20 per type (24 types, 10 same phoneme, 14 phoneme pairs three steps apart), 100 were null trials (silence), and 80 were control trials in which the phonemes in the pair had the same F1, F2, and F3 formants but differed in the fundamental frequency F0. Three fundamental frequencies were used: low, 108–8 1Hz; medium, 144–108 Hz; high, 192–144 Hz. The run orders were counterbalanced across subjects, and the event-related stimulus orderings were counterbalanced up to three back.

The subjects were given a simple alertness-maintaining task in the scanner that did not rely upon phonetic attention: they had to press a button when the second stimulus in the pair was quieter. The were 40 such quieter, button-press trials. The fMRI data from the differing-F0 trials, from the button-press trials, and from the false-alarm button-presses were excluded from the analysis.

### Behaviorally Weighted Contrast

We constructed a behaviorally weighted general linear model contrast, looking for brain areas that specifically amplify stimulus differences that are perceived as phonetic. The overall construction of this contrast is illustrated in Figure 3 in the main text. Specifically, for each subject the vector of same/different responses from the postscan behavioral test was zero-meaned to provide contrast coefficients to weight the fMRI data. For each three-step stimulus pair A and B, the fMRI term to be weighted was the activation corresponding to the following subtraction: (A&B + B&A) – (A&A + B&B). These behaviorally weighted fMRI terms were then summed to yield the overall contrast value for each voxel.

To express this in terms of an equation, let $f_{AB}$ represent the fMRI activation elicited by the stimulus pair A&B, and let $d_{AB}$ represent the proportion of times that the given subject perceived the pairs A&B and B&A as consisting of different phonemes, averaging across stimulus order. Let $m$ be the mean value of the $d_{AB}$ scores across all such A&B pairs. Then the overall contrast is equal to

$$\sum_{A,B}((d_{AB} - m) \times [(f_{AB} + f_{BA}) - (f_{AA} + f_{BB})]).$$

### Statistical Analysis

Preprocessing and statistical analysis of the data were performed using SPM99 software (Wellcome Dept of Cognitive Neurology). Preprocessing included slice timing correction, motion correction, normalization to the MNI305 stereotactic space (using linear and nonlinear basis functions, interpolating to 3 mm cubic voxels), and spatial smoothing with an 8 mm isotropic Gaussian kernel. Statistical analysis was performed using the general linear model. Global signal scaling was not applied, in order to prevent spurious deactivations. The event-related design was modeled using a canonical haemodynamic response and its temporal derivative. Comparisons of interest were implemented as linear contrasts. The subjects' individual same/different /ba/-/da/ discrimination scores were used to make perceptually weighted contrasts, after first being normalized and zero-meaned. This analysis was performed individually for each subject, and contrast images for each subject were used in a second-level analysis treating subjects as a random effect. For the functional connectivity analysis, the time course of the activity in the left supramarginal gyrus ROI was extracted and used as a regressor for each subject individually, looking across the whole brain for voxels whose time courses were correlated with it. This generated a set of 12 supramarginal-correlation contrast images, one from each subject, which were then passed into a standard second-level random-effects analysis. The random effects analyses were thresholded at the voxel level at $p < 0.001$ uncorrected and then corrected for multiple comparisons at the cluster level, at $p < 0.05$ (Friston et al., 1996). The statistical overlay images were made using MRIcro (Rorden and Brett, 2000).

### Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/56/4/726/DC1/.

## ACKNOWLEDGMENTS

The authors would like to thank Colin Phillips for having provided the SenSyn parameter files used for generating the synthetic speech stimuli and Alec Marantz for helpful early guidance. We would also like to thank Pat Kuhl, Sarah Hawkins, and Steve Stufflebeam for very helpful comments on various stages of the manuscript and Shanti Czaja for assistance with collecting the data. The fMRI scans were carried out at the Martinos Center/MGH-NMR Center, Charlestown, Massachusetts. R.D.S.R. was supported in part by a McDonnell-Pew Fellowship for Cognitive Neuroscience.

Received: September 27, 2006
Revised: July 13, 2007
Accepted: November 2, 2007
Published: November 20, 2007

## REFERENCES

Benedict, R.H., Shucard, D.W., Santa Maria, M.P., Shucard, J.L., Abara, J.P., Coad, M.L., Wack, D., Sawusch, J., and Lockwood, A. (2002). Covert auditory attention generates activation in the rostral/dorsal anterior cingulate cortex. J. Cogn. Neurosci. *14*, 637–645.

Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., and Liberman, A.M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. *78*, 364–396.

Binder, J., Liebenthal, E., Possing, E., Medler, D., and Ward, B. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. *7*, 295–301.

Blumstein, S., Myers, E., and Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. J. Cogn. Neurosci. *17*, 1353–1366.

Bornstein, M.H., Kessen, W., and Weiskopf, S. (1976). The categories of hue in infancy. Science *191*, 201–202.

Boynton, G., and Finney, E. (2003). Orientation-specific adaptation in human visual cortex. J. Neurosci. *23*, 8781–8787.

Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.

Brosch, M., and Schreiner, C. (1997). Time course of forward masking tuning curves in cat primary auditory cortex. J. Neurophysiol. *77*, 923–943.

Callan, D.E., Tajima, K., Callan, A.M., Kubo, R., Masaki, S., and 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.

Caplan, D., Gow, D., and Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. Neurology *45*, 293–298.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J., Berry, I., Nespoulous, J., and 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.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. Neuroreport *8*, 919–924.

Etcoff, N.L., and Magee, J.J. (1992). Categorical perception of facial expressions. Cognition *44*, 227–240.

Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science *291*, 312–316.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., and Frith, C.D. (1996). Detecting activations in PET and fMRI: levels of inference and power. Neuroimage *4*, 223–235.

Gelfand, S.A. (2004). Hearing: An Introduction to Psychological and Physiological Acoustics, Fourth Edition (New York: Informa Healthcare).

Gelfand, J.R., and Bookheimer, S.Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron *38*, 831–842.

Gerrits, E., and Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. Percept. Psychophys. *66*, 363–376.

Griffiths, T., Warren, J., Scott, S., Nelken, I., and King, A. (2004). Cortical processing of complex sound: a way forward? Trends Neurosci. *27*, 181–185.

Grill-Spector, K., and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. (Amst.) *107*, 293–321.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. *4*, 131–138.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J. Neurosci. *23*, 9541–9546.

Jiang, X., Rosen, E., Zeffiro, T., Vanmeter, J., Blanz, V., and Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. Neuron *50*, 159–172.

Kuhl, P.K., and Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. J. Acoust. Soc. Am. *73*, 1003–1010.

Leopold, D., O'Toole, A., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. Nat. Neurosci. *4*, 89–94.

Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. *54*, 358–368.

Liebenthal, E., Binder, J., Spitzer, S., Possing, E., and Medler, D. (2005). Neural substrates of phonemic perception. Cereb. Cortex *15*, 1621–1631.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.

Mathiak, K., Hertrich, I., Grodd, W., and Ackermann, H. (2002). Cerebellum and speech perception: a functional magnetic resonance imaging study. J. Cogn. Neurosci. *14*, 902–912.

May, P., Tiitinen, H., Ilmoniemi, R., Nyman, G., Taylor, J., and Naatanen, R. (1999). Frequency change detection in human auditory cortex. J. Comput. Neurosci. *6*, 99–120.

McEvoy, L., Levanen, S., and Loveless, N. (1997). Temporal characteristics of auditory sensory memory: neuromagnetic evidence. Psychophysiology *34*, 308–316.

Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature *385*, 432–434.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cogn. Neurosci. *12*, 1038–1055.

Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. Behav. Neurol. *12*, 191–200.

Rotshtein, P., Henson, R., Treves, A., Driver, J., and Dolan, R. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat. Neurosci. *8*, 107–113.

Smiarowski, R.A., and Carhart, R. (1975). Relations among temporal resolution, forward masking, and simultaneous masking. J. Acoust. Soc. Am. *57*, 1169–1174.

Wilson, R.H., and Carhart, R. (1971). Forward and backward masking: interactions and additivity. J. Acoust. Soc. Am. *49* (*Suppl 2*), 1254.

Zevin, J.D., and McCandliss, B.D. (2005). Dishabituation of the bold response to speech sounds. Behav. Brain Funct. *1*, 4.