Short-Term Reorganization of Auditory Analysis Induced by Phonetic Experience

Einat Liebenthal[1], Jeffrey R. Binder[1], Rebecca L. Piorkowski[2], and Robert E. Remez[2]

Abstract

■ Sine wave replicas of spoken words can be perceived both as nonphonetic auditory forms and as words, depending on a listener's experience. In this study, brain areas activated by sine wave words were studied with fMRI in two conditions: when subjects perceived the sounds spontaneously as nonphonetic auditory forms ("naïve condition") and after instruction and brief practice attending to their phonetic attributes ("informed condition"). The test items were composed such that half replicated natural words ("phonetic items") and the other half did not, because the tone analogs of the first and third formants had been temporally reversed ("nonphonetic items"). Subjects were asked to decide whether an isolated tone analog of the second formant (T2) presented before the sine wave word (T1234) was included in it. Experience in attending to the phonetic properties of the sinusoids interfered with this auditory matching task and was accompanied by a decrease in auditory cortex activation with word replicas but not with the acoustically matched nonphonetic items. Because the activation patterns elicited by equivalent acoustic test items depended on a listener's awareness of their phonetic potential, this indicates that the analysis of speech sounds in the auditory cortex is distinct from the simple resolution of auditory form, and is not a mere consequence of acoustic complexity. Because arbitrary acoustic patterns did not evoke the response observed for phonetic patterns, these findings suggest that the perception of speech is contingent on the presence of familiar patterns of spectral variation. The results are consistent with a short-term functional reorganization of auditory analysis induced by phonetic experience with sine wave replicas and contingent on the dynamic acoustic structure of speech. ■

INTRODUCTION

What is the role of auditory analysis in phonetic perception? Is a phonetic percept determined solely by the auditory properties of speech, or is it contingent on perception in a phonetic mode, which is engaged when sounds are recognized as speech and which is distinct from the auditory (nonspeech) mode? Liberman and colleagues (Liberman et al., 1967; Liberman & Mattingly, 1989) maintain that speech perception is embodied in the brain in a module devoted to phonetic analysis. In this account, speech sounds are inevitably analyzed by a dedicated function that reciprocates the articulatory gestures of vocal sound production. In contrast, others (e.g., Bregman, 1990; Kluender & Greenberg, 1989) have argued that phonetic perception is governed by auditory analysis indifferent to the vocal origin of sound. In this concept, the phonetic percept evoked by a speech sample emerges from its auditory analysis.

These questions have been pivotal in forming contemporary accounts of speech perception (Bregman, 1990; Kluender & Greenberg, 1989; Liberman & Mattingly, 1989). To provide part of the answer, several functional imaging studies using positron emission tomography or functional magnetic resonance imaging (fMRI) have focused on the physiological correlates of auditory and phonetic perceptual processes. By comparing brain regions that are activated by speech and nonspeech sounds, these studies implicated several nonprimary auditory areas within the superior temporal cortex in the analysis of sounds (Vouloumanos, Kiehl, Werker, & Liddle, 2001; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Binder, Frost, et al., 2000; Scott, Blank, Rosen, & Wise, 2000; Mummery, Ashburner, Scott, & Wise, 1999; Demonet et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992). However, these studies elicited auditory and phonetic perception with unique sets of acoustic materials aimed to evoke each function. This method confounds the contrast between phonetic and nonphonetic perception with the auditory effects of the phonetic and nonphonetic test items.

In the present experiment, our goal was to study the effect of the mode of perception, auditory or phonetic, in the analysis of speech sounds, independent of the acoustic properties of the sounds. To this purpose, we used a single set of sine wave replicas of spoken words, which can be perceived both as nonphonetic auditory forms and as words, depending on a listener's experience (Best, Studdert-Kennedy, Manuel, & Rubin-Spitz, 1989; Remez, Rubin, Pisoni, & Carrell, 1981). In this kind of

[1]Medical College of Wisconsin, [2]Barnard College

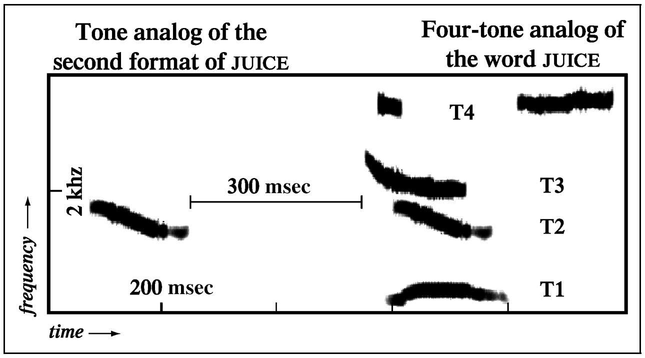

Figure 1. Spectrogram of the isolated tone analog of the second formant (T2) of the word JUICE (left) and the four-tone replica (T1234) of the word JUICE (right).

synthetic acoustic signal, three or four time-varying sinusoids replicate the frequency and amplitude variation of the oral, nasal, and fricative resonances—the formants—of a speech sample. A sine wave replica of an utterance is perceptually bistable (i.e., it can evoke both nonphonetic auditory forms [patterns of changing pitch and loudness] and phonetic forms [a series of consonants and vowels]) because it preserves the dynamic acoustic properties of natural speech, evoking phonetic perception while lacking harmonic structure, momentary aperiodicities, and regular pulsing, the fine-grain acoustic properties of speech. A listener who is unfamiliar with this kind of signal typically resolves the time-varying sinusoids into a complex of contrapuntal tones varying concurrently in pitch and loudness, without phonetic properties. However, when informed that the tones compose speech, a listener is able to identify the linguistic properties after very brief practice. This perceptual bistability affords a test of the functional difference between auditory perception and phonetic perception without varying the test material between conditions.

In composing a fair test with the potential to distinguish auditory perception from phonetic perception, we employed an auditory task that could be performed with or without knowledge of the phonetic properties of the test items. Because this task required a subject to resolve individual tones of a sine wave replica, we supposed that phonetic attention to the tone complexes might interfere with the task relative to the condition in which the subject was not aware of the phonetic attributes of sine wave speech. In addition to the main question of the physiological correlates of phonetic perception, we assessed the role of the auditory forms initiating these processes. Would an arbitrary, albeit complex, sine wave pattern elicit a phonetic perceptual state in the listener after experience with phonetically effective sine wave patterns? This question was addressed by inclusion of test items that closely resembled the sine wave replicas of utterances but which had been altered acoustically to preclude phonetic perception.

Brain areas activated by sine wave words were studied by monitoring the blood oxygenation level-dependent (BOLD) response with fMRI in two conditions: when subjects perceived the sounds without instruction or training as to their phonetic properties ("naïve condition") and after instruction and brief practice attending to their phonetic attributes ("informed condition"). The test items were composed such that half replicated natural words ("phonetic items") and the other half did not, because the tone analogs of the first and third formants had been temporally reversed ("nonphonetic items"). Subjects were asked to decide whether an isolated tone analog of the second formant (T2), presented before the sine wave complex (T1234), was included in it (Figure 1; Remez, Pardo, Piorkowski, & Rubin, 2001). Between the naïve and the informed conditions, subjects were informed that the test items were derived from speech, and they practiced transcribing sine wave sentences and then single words in two four-alternative, forced-choice (4-AFC) tasks. The experimental paradigm is described in Table 1. The only difference between the naïve and informed conditions was the listener's awareness of the phonetic potential of the sine wave words. In neither condition

Table 1. Test Blocks and Tasks

Test Block	*Condition*	*Sinusoidal Test Items*	*Task*
	Auditory form practice	Tones	1. T2–T2 matching
			2. T2–T1234 matching
1	Naïve 1	Phonetic/nonphonetic word-length items	T2–T1234 matching
2	Naïve 2	Phonetic/nonphonetic word-length items	T2–T1234 matching
	Phonetic practice	1. Sentences	4-AFC transcription
		2. Words	4-AFC transcription
3	Informed 1	Phonetic/nonphonetic word-length items	T2–T1234 matching
4	Informed 2	Phonetic/nonphonetic word-length items	T2–T1234 matching

See the Experimental Procedure under Methods for details.

were subjects instructed that only half of the tone complexes were phonetically veridical, replicating words, while the other half were arbitrary constructions and, therefore, nonphonetic.

RESULTS

Behavioral and neural measures alike revealed two effects of listening to the tone patterns. First, a phonetic item elicited a different behavioral response and functional activation pattern when it was perceived as a nonphonetic auditory form and when it was perceived as a sound with phonetic potential. Second, in the informed condition, arbitrary tone patterns without phonetic potential evoked a different behavioral response and activation pattern than sine wave replicas of words. The fact that the activation patterns elicited by equivalent acoustic test items depended on a listener's awareness of their phonetic potential indicates that the analysis of speech sounds performed in the auditory cortex involves more than the simple resolution of auditory form and is not a mere consequence of acoustic complexity. Because arbitrary acoustic patterns did not evoke the behavioral change or change in neural activation observed for phonetic patterns, these findings suggest that the processes leading to perception of speech are contingent on both the awareness of phonetic potential and the presence of spectral variation consistent with speech.

Perceptual Measures

The effects of test block on performance accuracy and on response time (RT) were studied separately for the phonetic and the nonphonetic items because it was hypothesized that the effects would differ (Figure 2). Specifically, we hypothesized that phonetic practice might interfere with performance particularly with the phonetic items. In contrast, proficiency with the task in the naïve state should not differ between the two types of test items. Overall, we observed three effects on task performance over the course of the experiment: 1) practice recognizing the phonetic content[1] of sine wave words interfered with later performance of the auditory matching task though only for phonetic test items; 2) proficiency with the task increased for both phonetic and nonphonetic test items between the first and second naïve test blocks; and 3) for the phonetic items, performance returned to its level before phonetic practice in block Informed 2. For the nonphonetic items, performance remained stable in the informed blocks. These conclusions are based on significant changes in RT with test blocks.

For the phonetic items, there was a significant effect of test block on RT, analysis of variance (ANOVA), $F(3,120) = 2.96$, $p < .04$, although this was not accompanied by a parallel effect of performance accuracy,

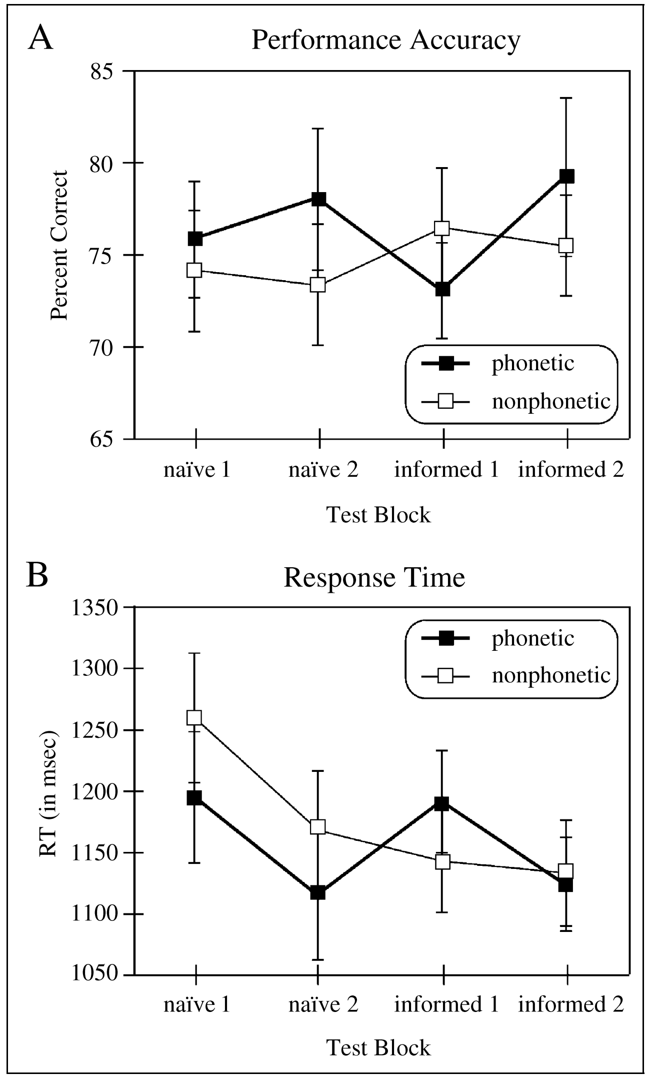

Figure 2. (A) Average group-performance accuracy (percent correct responses) and (B) average group-response latency for the auditory matching task in each condition. Error bars indicate 95% confidence intervals. RTs were measured from the onset of the tone complexes.

ANOVA, $F(3,120) = 2.05$, $p < .11$. This effect was attributable to a decrease in RT from 1195 to 1115 msec between blocks Naïve 1 and Naïve 2 ($p < .02$), an increase to 1191 msec after phonetic practice ($p < .03$) and a decrease to 1124 msec in block informed 2 ($p < .05$). For the nonphonetic items, performance accuracy did not change significantly across blocks, ANOVA, $F(3,120) = .74$, $p < .53$, but RT decreased, ANOVA, $F(3,120) = 5.61$, $p < .001$. RT decreased from 1259 to 1168 msec between blocks Naïve 1 and Naïve 2 ($p < .01$) and did not change significantly subsequently.

Reliable differences in RT but not in accuracy were observed between the phonetic and nonphonetic test items during the naïve condition, RT: ANOVA, $F(1,122) = 4.44$, $p < .04$; accuracy: ANOVA, $F(1,122) = 3.12$, $p < .07$.

Performance in the phonetic form practice with sentences averaged 55 ± 26% correct in the first cycle of the task and improved to 84 ± 21% correct by the fourth cycle. In the phonetic form practice with words, performance averaged 60 ± 16% correct (chance = 25% in both tasks). These scores indicate that the subjects were able to perceive the correct phonetic form of the sine wave items well above chance but not perfectly.

In the subjective report at the end of the session, 29 of 31 subjects confirmed being unaware of the phonetic potential of the test items during the naïve condition of the experiment; 13 of 31 subjects confirmed recognizing words during the informed condition of the task.

Physiological Measures

To identify the neural substrates for auditory and phonetic listening, we contrasted the activation in the informed test blocks (3 and 4) with activation in the naïve test blocks (1 and 2).[2] As a control for proficiency effects, activation in the later test block in each condition (Naïve 2 and Informed 2) was contrasted with activation in the earlier test block in each of the conditions (Naïve 1 and Informed 1). Because behavioral performance revealed differential effects of phonetic practice on the response to phonetic and nonphonetic test items, functional activation evoked by the two types of test items was analyzed separately.

Several foci exhibited decreased activation relative to the naïve condition following phonetic practice: in the temporal lobes, in the left frontal lobe and in several medial structures. No areas exhibited increased activation because of phonetic practice (Figure 3 and Table 2). The largest and most significant focus of decreased activation for the phonetic items was in the left superior temporal lobe, with peaks in the left anterolateral part of Heschl's gyrus (HG) and in the posterior superior temporal gyrus (STG), corresponding to secondary and association auditory cortex (Howard et al., 2000; Liegeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994; Galaburda & Sanides, 1980; Celesia, 1976). The right lateral HG also exhibited this differential activation, although less extensively. With the nonphonetic items, a single focus was observed in the left posterior STG that was much less extensive than with the phonetic items. Differential activation in these temporal areas is probably due specifically to phonetic practice and not general familiarity with the task because this pattern was not seen in the proficiency contrasts. The dorsomedial thalamic nucleus and medial prefrontal Brodmann's area (BA) 8 also exhibited this pattern of decreased activation for the phonetic items following phonetic training. In the control (proficiency) contrasts, a pattern of decreasing activation with later test blocks was found in the left prefrontal cortex (BA 9 and 44) for both phonetic and nonphonetic items, indicating that these areas may be less active when proficiency is gained with the task.

It is noteworthy that for all the activation foci discussed above (summarized in Table 2), the average BOLD signal evoked during the task condition exceeded the average signal evoked during the silent baseline intervals. That is, the decrease in activation observed

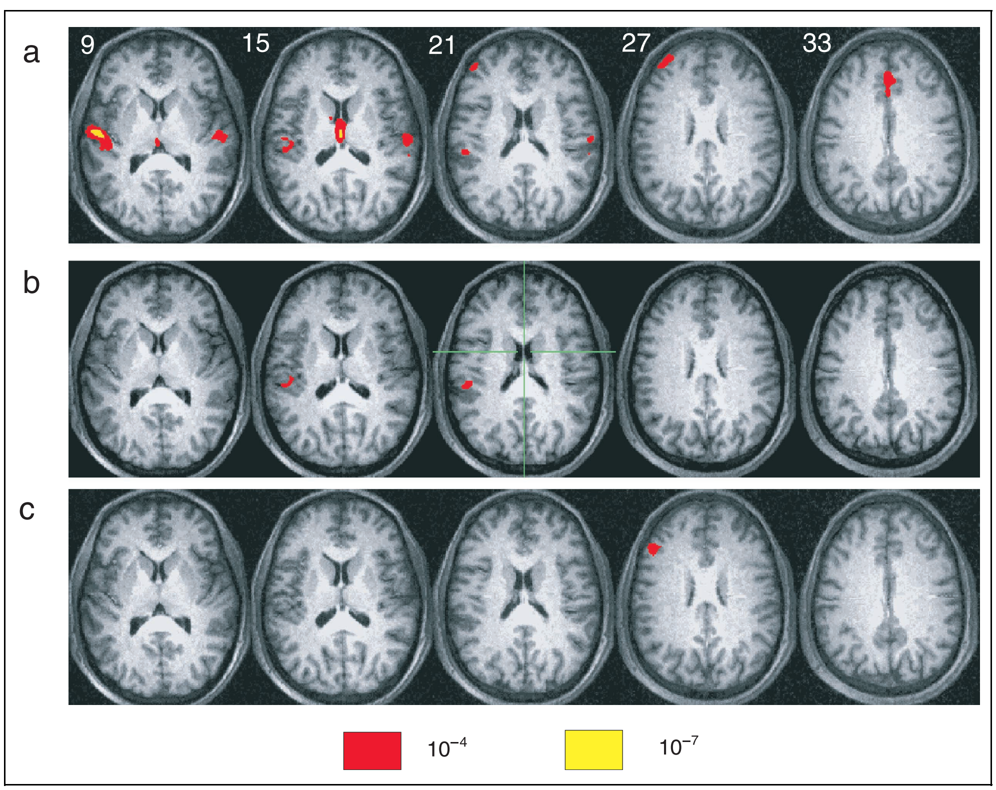

Figure 3. Statistical parametric maps of the informed-naïve contrast with (a) phonetic and (b) nonphonetic test items, and (c) of the proficiency contrast with phonetic items, collapsed across subjects. The stereotaxic *z* coordinate of each axial section is given on the top image. Color coding indicates uncorrected voxel-wise *p* values. The green crosshairs indicate the AC–PC line (vertical) and the AC plane (horizontal). (a, b) Highlighted foci indicate voxel clusters with significant "decrease" in activation in the informed condition (test blocks 3 and 4) relative to the naïve condition (test blocks 1 and 2). (c) Highlighted foci indicate voxel clusters with significant decrease in activation in the later test blocks (Naïve 2 and Informed 2) relative to the earlier test blocks (Naïve 1 and Informed 1) within each condition.

Table 2. Deactivation Peaks

Talairach Coordinates (mm)						
x	*y*	*z*	*t Value*	*Cluster Size (μl)*	*Anatomical Location*	*BA*
Informed-Naïve Contrast, Phonetic						
−55	−13	9	−0.61	2844	Left HG	42
−51	−30	19	−0.43	2844	Left posterior STG	42/22
58	−19	14	−0.50	1645	Right HG	42
−1	−16	14	−0.58	1413	Dorsomedial nucleus of thalamus	
0	35	33	−0.49	1038	Superior frontal gyrus	8
−34	53	26	−0.45	730	Left middle frontal gyrus	10
Informed-Naïve Contrast, Nonphonetic						
−50	−30	19	−0.43	657	Left posterior STG	42/22
Proficiency Contrast, Phonetic						
−48	28	27	−0.44	711	Left middle frontal gyrus	9
Proficiency Contrast, Nonphonetic						
−56	12	31	−0.43	650	Left inferior frontal gyrus	44

Coordinates in standard stereotaxic space (Talairach & Tournoux, 1988), approximate anatomical locations and Brodmann's areas (BA), corresponding t values of deactivation peaks, and cluster sizes of deactivation foci in clusters larger than 300 μl and with t values above 0. 37 (uncorrected voxel-wise $p < 10^{-4}$), for all condition contrasts. x = lateral distance from anterior commissure–posterior commissure (AC–PC) line (positive right); y = anterior–posterior distance from AC (positive is in front of AC); z = distance above–below AC–PC line (positive is above).

in the informed relative to the naïve condition was because of a true decrease in activation relative to baseline in these foci during the informed condition.

DISCUSSION

Phonetic and Auditory Perception

Is phonetic perception different from general auditory perception? In our study, listeners treated tone analogs of speech differently in the naïve and informed conditions, as revealed by measures of both perceptual performance and physiology, despite the absence of any change in acoustic structure. Phonetic awareness interfered with performance on the auditory matching task, although only in trials when the tone complex replicated actual phonetic spectrotemporal changes. The matching task required a listener to remember the auditory form of the single tone presented in the first interval and to compare it to the pattern of tones that followed. To succeed in this task, a listener must resolve the sine wave word into its constituent tones, consistent with contemporary descriptions of auditory perceptual organization (Bregman, 1990). However, phonetic perceptual organization opposes this resolution because it requires fusion of the tones into a single auditory object to analyze phonetically (Remez, Rubin, Berns, Pardo, & Lang, 1994).

A direct measure of phonetic perception could not be obtained because the objectives of this study required that the same (auditory) task be used in the naïve and in the informed conditions. However, several indirect measures indicate that subjects perceived almost exclusively the auditory form of the sine wave words in the naïve condition, whereas they perceived both the auditory and the phonetic forms after phonetic practice. First, the subjective reports confirm that subjects were largely unaware of the phonetic potential of the test items during the naïve condition. Second, performance in the phonetic form practice tasks indicates that subjects were able to perceive the phonetic form of the sine wave words well above chance, when concentrating on this form alone. Third, RT measures confirm that phonetic practice interfered with auditory analysis of the phonetic test items. Fourth, in their subjective report, fewer than 50% of the subjects confirmed recognizing words during the informed condition of the task. These measures suggest that subjects did not focus on the phonetic form of the sine wave words during the informed condition, probably because it was irrelevant to (and interfered with) the auditory matching task. Therefore, it is plausible to conclude that subjects attended solely to the auditory form of the sine wave words in the naïve condition. In the informed condition, they might have resolved the phonetic form of the words to some extent while essentially focusing on the auditory form of the words.[3]

Differences in RT were also observed between the phonetic and the nonphonetic test items during the naïve condition. These differences were not expected. Inspection of the behavioral response by test item and of the acoustic structure of the test items did not reveal systematic effects that could account for this observation.

Empirical precedent has shown that phonetic organization of time-varying sinusoids requires three conditions. First, the variation of the tones must replicate the spectrotemporal properties of actual speech. Arbitrary patterns of tone variation are phonetically ineffective despite similar distributions of frequency variation (Remez et al., 1994). Second, phonetic perception does not ensue unless the tone analogs of the first and second formant are available concurrently (Remez et al., 1981). Together, these two observations suggest that spectrotemporal patterns created by the combination of first and second formants in natural speech define a subset of such patterns that can, as a result of either experience or innate ability, be perceived as speech phonemes. These conditions are not satisfied by the nonphonetic items of our test. Although the range and rate of frequency variation of each constituent is derived from natural speech, the arbitrary temporal reflection of one constituent disrupts the original temporal relation between the first and second tones. Third, the listener must intend to hear the tones phonetically despite their anomalous auditory quality (Best et al., 1989; Remez et al., 1981). Although the demanding auditory matching task used throughout the present experiment evidently prevented subjects from hearing the tones phonetically, brief practice with sine wave words between Blocks 2 and 3 appears to have induced a short-term reorganization of auditory analysis that promoted the fusion of tones when the pattern of variation was consistent with a phonologically governed vocal origin. The task performance and BOLD evidence confirm the auditory functional differentiation of trials in which phonetic fusion was possible from trials in which phonetic fusion was impossible. This suggests that an auditory function sensitive to dynamic attributes of vocal sounds is central to the perceptual organization of speech.

Reorganization of Auditory Cortex Analysis

Auditory cortical areas were activated by a sinusoidal pattern with phonetic potential, and the listener's susceptibility to such sounds as speech resulted in decreased activation in those areas, though only when the sinusoids were consistent with speech. The results are consistent with a short-term functional reorganization of auditory analysis induced by phonetic experience with sine wave replicas and contingent on the dynamic acoustic structure of speech.

Partial corroboration for the involvement of the STG in the auditory matching task can be found in several animal studies. Secondary and association auditory fields in the STG in primates have been implicated in the analysis of complex sounds (Kosaki, Hashikawa, He, & Jones, 1997; Rauschecker, Tian, & Hauser, 1995) and in auditory short-term memory (Colombo, D'Amato, Rodman, & Gross, 1990). The auditory cortex deactivation probably reflects decreased execution, following phonetic practice, of those processes required for analytic listening in the matching task, and this deactivation underlies the reduced performance level. The interfering attention to the fused attributes of the tone complexes can be compared to an increase in noise causing decrease in auditory salience. The phonetic interference reported here, like background noise, acts as a mask of relevant auditory information. Although the physiological mechanisms underlying these effects may differ, it is noteworthy that declines in the discriminability of speech in noise have been previously associated with decreases in BOLD activation in anterolateral HG bilaterally (Binder, Possing, et al., 2000; Colder & Tanenbaum, 1999; Shah, Jancke, Grosse-Ruyken, & Muller-Gartner, 1999). The left-lateralized locus of differential activation in HG observed in this study can be related to a more extensive reorganization process induced in that hemisphere by phonetic practice. This is consistent with a specialization of this hemisphere for perceptual organization of speech.

The dorsomedial thalamic nucleus and medial prefrontal BA 8 also exhibited this pattern of decreased activation for the phonetic items following phonetic training, although the functional significance of this observation is unknown. In the macaque, the dorsomedial thalamic nucleus has reciprocal connections with auditory fields predominantly in the posterior STG (Pandya, Rosene, & Doolittle, 1994). The medial prefrontal cortex has reciprocal connections with auditory fields in the STG (Barbas, Ghashghaei, Dombrowsky, & Rempel-Clower, 1999; Petrides & Pandya, 1988) and also receives thalamic projections from the dorsomedial nucleus (Giguere & Goldman-Rakic, 1988). Thus, the auditory association cortex, dorsomedial thalamus, and medial prefrontal cortex may be components of a neural system engaged by the auditory matching task that are susceptible to interference by concurrent perceptual fusion of the tones.

The left inferior frontal (BA 44) and middle frontal (BA 9) cortex showed decreased activation with an increase in task proficiency. These areas may be recruited when learning a new and demanding task and then become progressively less activated when skilled performance is achieved. The left prefrontal cortex has been shown to be part of a network recruited during initial, effortful performance of a verbal production task

(Raichle et al., 1994). The activation associated with the naïve-informed contrast did not overlap with the pattern of activation in the proficiency contrast, confirming that the naïve-informed measures were not attributable to the effects of changing proficiency.

Although the auditory task employed here did not require attention to phonetic detail, it is noteworthy that the auditory cortex deactivation was not accompanied by a corresponding shift of activity to other areas concerned with phonetic perception. Given that subjects were engaged in a demanding auditory matching task, it is unlikely that conscious phonetic perception occurred during scanning. This claim is supported by the fact that no subject reported awareness of the presence of nonphonetic items. These observations indicate that, at least with sine wave signals, the early auditory perceptual organization of speech can be independent of the phonetic perception of speech.

Conclusions

This study focused on factors affecting the early auditory analysis of speech sounds. The dual perceptual properties of the sine wave replicas allowed a clear distinction to be made between the effects of perceptual experience and the effects of acoustic structure. The extraordinary aspect of the auditory organization of speech perception revealed here is that it was induced by phonetic exposure, but its incidence was entirely determined by the immediate acoustic structure of the sounds. Whether this reflects a general auditory susceptibility to the spectrotemporal structure of speech, an adaptation of the general auditory system resulting from lifelong exposure to these spectrotemporal patterns, or a function dedicated to phonetic organization remains uncertain. Nonetheless, the short-term reorganization of auditory analysis, contingent both on experience with the phonetic attributes of tone analogs of speech and on the structure of sound itself, may reflect the means by which speech is differentiated from other auditory events.

METHODS

Subjects

Results are reported from 31 subjects (18 men) who responded on at least 2/3 of the trials in every test block and performed above chance level (above 58.3%, chi-square test, $p < .05$) in the average of all test blocks. Subjects were right-handed, native speakers of English, aged 18–57 years.

Test Items

The sine wave words were prepared according to procedures detailed by Remez, Rubin, Nygaard, and Howell (1987) and Remez et al. (2001). The nonphonetic complexes were prepared by temporally reversing their first and third formant tone analogs. The test items were drawn from a pool of 108 different words, each presented in phonetic and matched nonphonetic form. One third of the items were repeated one time during the experiment. Half of the trials involved phonetic items and half involved matched nonphonetic items. Half of the trials were "false" trials in which the isolated tone was the analog of the second formant of a different tone complex that differed from the "true" complex by a single vowel step when ordering the sine wave vowels by the average frequency of the second formant at the vowel nucleus.

The intelligibility of the phonetic items was evaluated in two ways, in a pilot study with 52 listeners. A free transcription test was conducted, followed by a three-alternative, forced-choice procedure in which subjects were asked to identify the sine wave word from three printed alternatives. The foils differed from the target by a single, distinctive phonetic feature in an initial or a final consonant. Mean performance was 52.1% correct and 89.5% correct, respectively, for the two tests. Similar tests could not be conducted with the nonphonetic items because they could not be transcribed.

A second pilot study with 18 subjects and a task consisting of rating the intelligibility of each item (phonetic and nonphonetic) along a scale of 1–3 (1, *clearly identifiable word*; 2, *word, but not clearly identifiable*, 3, *nonspeech*) showed that 61% of the phonetic items were considered clearly identifiable words (with 22% considered nonspeech), whereas 58% of the nonphonetic items were considered nonspeech (with 20% considered identifiable words).

Experimental Procedures

The test blocks and tasks are summarized in Table 1.

Before entering the scanner, each subject practiced the auditory matching task (auditory form practice), consisting initially of determining whether two successive single sinusoids were identical in pitch contour. Ten trials were presented, and 90% correct performance was required. All subjects reached the required level of performance after one cycle of the task except for one subject who had to repeat the task twice to reach the criterion. Subjects then practiced determining whether a single sinusoidal tone presented in the first interval (always the second formant, or T2 tone of a complex) was a constituent of a complex of three or four tones presented after a 300-msec silence in the second interval (T2–T1234 matching task). (See Figure 1 for a description of this task.) Blocks of 10 trials were presented until 90% correct performance was reached in any one block or 75% correct in two consecutive blocks. All subjects reached the required level of performance after five blocks or less. The material used for auditory form practice consisted of arbitrarily composed sinusoids devoid of phonetic content.

In the scanner, the T2–T1234 auditory matching task was presented in two conditions. In the naïve condition (test blocks 1–2), the test items were referred to as "single tone" and "tone complex." After test block 2, scanning was halted and the subject was informed that the tone complexes were derived from English words. To provide brief practice in sine wave speech recognition, eight sentences and then 18 isolated words not used during scanning, each realized as sine wave patterns, were presented in a four-alternative, forced-choice (4-AFC) task, in which four possible transcriptions of the sine wave sentence or word were displayed on a screen after the subject had heard the item. The sentences task was cycled four times and on a fifth cycle the correct transcription was displayed. The word task was presented only once. Subjects' performance was not scored online and no performance criteria were set for proceeding with the experiment. Scanning then resumed with two additional test blocks (test blocks 3–4) of the T2–T1234 auditory matching task (informed condition). In the informed condition, the tone complexes were referred to as "words" (although half of the trials involved nonphonetic items, as in the naïve condition). At the end of the experiment, subjects were asked to say whether the tone patterns had evoked impressions of words during testing and whether they were aware that some of the items were not words.

Image Acquisition and Analysis

Images were acquired on a 3T Bruker scanner. Functional data consisted of T2*-weighted, gradient-echo, echo-planar images (TE = 27.2 msec, flip angle = 90°, NEX = 1), obtained using clustered acquisition (acquisition time = 1190 msec) at 9-sec intervals to avoid perceptual masking of the test items or contamination of the data by the acoustic noise of the scanner (Edmister, Talavage, Ledden, & Weisskoff, 1999). Two trials were presented within each interval between image acquisitions, both consisting of either phonetic or nonphonetic test items. The order of presentation of phonetic and nonphonetic trials within a test block was randomized, except that every third trial consisted of a silent baseline with no acoustic presentation or task. In each test block, 18 images per trial type (phonetic, nonphonetic, baseline) were acquired. There were two test blocks per condition (Naïve 1 and Naïve 2, Informed 1 and Informed 2) such that a total of 36 images were acquired per experimental condition and trial type. The images were reconstructed from 16 axially oriented contiguous slices with 3 × 3 × 4 mm voxel dimensions. Slice coverage was centered obliquely around the temporal lobes and varied by a few millimeters from subject to subject, such that the brain tissue between stereotaxic z coordinates 2 and 38 in the anterior part of the brain and between stereotaxic z coordinates −12 and 22 in the posterior part of the brain was imaged in all 31 subjects. This volume includes most of the temporal lobes (except for the anterior most ventral tip of the superior, middle and inferior temporal gyri), part of the frontal and parietal lobes (including all the inferior frontal gyrus and missing the dorsal portions of the middle and superior frontal gyri and the dorsal portion of the parietal lobe) and the occipital lobe. High-resolution anatomical images of the entire brain were obtained using a 3-D spoiled-inversion-recovery sequence, with 0.9 × 0.9 × 1.2 mm voxel dimensions.

Image analysis was done with the AFNI software package (Cox, 1996). Individual statistical parametric maps were created as follows: Images acquired after baseline trials were subtracted from the task images acquired immediately preceding and immediately following them, to create BOLD difference maps or activation maps. This use of a local baseline reduces effects of signal instability caused by scanner drifts and low-frequency spontaneous fluctuations. The difference maps were sorted according to trial type and experimental condition and were subjected to a voxel-wise analysis of variance (ANOVA) to detect the effects of experimental condition (naïve, informed) and test block (first or second within a condition) on the pattern of activation evoked by each type of test item. The difference maps were also simply averaged according to experimental condition to assess the BOLD signal relative to baseline in each condition. Individual anatomical scans and statistical maps were projected into standard stereotaxic space (Talairach & Tournoux, 1988). The statistical maps were smoothed with a Gaussian filter measuring 6 mm full width half maximum to compensate for individual variation in anatomy across subjects. The statistical maps were merged across subjects by averaging the t values at each voxel (Binder, Frost, et al., 2000; Binder et al., 1997). The procedure of averaging t statistics was adopted to guard against heteroscedasticity of the MR signal variance among subjects, which can be caused, for instance, by differing degrees of subject motion or tissue pulsatility, variability in global blood flow or reactivity, or scanner variability between sessions. Randomization testing was used to determine the threshold t values for rejecting the null hypothesis (Bullmore et al., 1996). Average t values of 0.37 or larger were considered significant (uncorrected voxel-wise $p < 10^{-3}$). Activation foci smaller than 300 μl were removed, which lowered the probability of false positives to $p < 10^{-4}$ as determined by Monte Carlo simulation using the AlphaSim module in AFNI (Ward, 2000).

Acknowledgments

We thank L. E. Bernstein, C. A. Fowler, I. G. Mattingly, and P. E. Rubin for comments on the manuscript, and E. T. Possing, J. N. Kaufman, B. D. Ward, R. Tong, A. Jesmanowicz, C. Y. Yang, and S. Wissig for technical assistance. This research was supported by NIH grants DC00308 (to Barnard College) and NIH grants NS33576 and MH51358 (to Medical College of Wisconsin).

Reprint requests should be sent to Einat Liebenthal, Department of Neurology, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA, or via e-mail: einatl@mcw.edu.

The data reported in this experiment have been deposited in The fMRI Data Center (http://www.fmridc.org). The accession number is 2-2002-113B6.

Notes

1. It should be noted that the phonetic test items compose words, and contain additional levels of information beyond the phonetic, when they are accessed linguistically.
2. The limited number of images acquired per experimental condition and type of test item (36) did not allow searching for effects between test blocks, as performed with the behavioral data.
3. In a previous study (Remez et al., 2001), subjects reported both the auditory and the phonetic form of sine wave words well when attending to both attributes in a dual task.

REFERENCES

Barbas, H., Ghashghaei, H., Dombrowsky, S. M., & Rempel-Clower, N. L. (1999). Medial prefrontal cortices are unified by common connections with superior temporal cortices and distinguished by input from memory-related areas in the rhesus monkey. *Journal of Comparative Neurology, 410,* 343–367.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309–312.

Best, C. T., Studdert-Kennedy, M., Manuel, S., & Rubin-Spitz, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception and Psychophysics, 45,* 237–250.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10,* 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17,* 353–362.

Binder, J. R., Possing, E. T., Kaufman, J. N., Bellgowan, P. S. F., Hammeke, T. A., Tong, R., Cox, R. W., & Ward, B. D. (2000). Distinct brain systems for sensory and decision processes in speech perception. *Cognitive Neuroscience Society,* 3B.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge: MIT Press.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., & Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance Methods, 35,* 261–277.

Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain, 99,* 403–414.

Colder, B. W., & Tanenbaum, L. (1999). Dissociation of fMRI activation and awareness in auditory perception task. *Cognitive Brain Research, 8,* 177–184.

Colombo, M., D'Amato, M. R., Rodman, H. R., & Gross, C. G. (1990). Auditory association cortex lesions impair auditory short-term memory in monkeys. *Science, 247,* 336–338.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonances neuroimages. *Computational Biomedical Research, 29,* 162–173.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115,* 1753–1768.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping, 7,* 89–97.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190,* 597–610.

Giguere, M., & Goldman-Rakic, P. S. (1988). Mediodorsal nucleus: Areal, laminar, and tangential distribution of afferents and efferents in frontal lobe of rhesus monkeys. *Journal of Comparative Neurology, 277,* 195–213.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., & Brugge, J. F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *Journal of Comparative Neurology, 416,* 79–92.

Kluender, K. R., & Greenberg, S. (1989). A specialization for speech perception. *Science, 244,* 1530–1531.

Kosaki, H., Hashikawa, T., He, J., & Jones, E. G. (1997). Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *Journal of Comparative Neurology, 386,* 304–316.

Liberman, A. M., Cooper, F. S., Sankweiler, D. A., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–481.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489–494.

Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked-potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology, 92,* 204–214.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. S. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America, 106,* 449–457.

Pandya, D. N., Rosene, D. L., & Doolittle, A. M. (1994). Corticothalamic connections of auditory-related areas of the temporal lobe in the rhesus monkey. *Journal of Comparative Neurology, 345,* 447–471.

Petrides, M., & Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *Journal of Comparative Neurology, 273,* 52–66.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, M. A. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4,* 8–26.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science, 268,* 111–114.

Remez, R. E., Pardo, J. S., Piorkowski, P. E., & Rubin, P. E. (2001). On the perceptual bistability of sine wave analogues of speech. *Psychological Science, 12,* 24–29.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A.

(1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 30–61.
Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional cues. *Science, 212,* 947–950.
Scott, S. K., Blank, C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400–2406.
Shah, N. J., Jancke, L., Grosse-Ruyken, M., & Muller-Gartner, H. W. (1999). Influence of acoustic masking noise in FMRI of the auditory cortex during phonetic discrimination. *Journal of Magnetic Resonance Imaging, 9,* 19–25.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* New York: Thieme Medical Publishers.
Vouloumanos, A., Kiehl, J. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994–1005.
Ward, B. D. (2000). Simultaneous inference for fMRI data [On-line]. Available: http://afni.nimh.nih.gov/afni/docpdf/.
Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and speech discrimination in speech processing. *Science, 256,* 846–849.