Behavioral/Systems/Cognitive

Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing

Kristin McNealy,[1,5,6] **John C. Mazziotta,**[1,3,4,5] **and Mirella Dapretto**[1,2,3,5,6]

[1]Ahmanson-Lovelace Brain Mapping Center, Semel Institute for Neuroscience and Human Behavior, [2]Department of Psychiatry and Biobehavioral Sciences, [3]The Brain Research Institute, [4]Departments of Neurology, Pharmacology, and Radiological Sciences in the David Geffen School of Medicine, [5]Neuroscience Interdepartmental Program, and [6]Center for Culture, Brain, and Development, University of California, Los Angeles, Los Angeles, California 90095

Word segmentation, detecting word boundaries in continuous speech, is a critical aspect of language learning. Previous research in infants and adults demonstrated that a stream of speech can be readily segmented based solely on the statistical and speech cues afforded by the input. Using functional magnetic resonance imaging (fMRI), the neural substrate of word segmentation was examined on-line as participants listened to three streams of concatenated syllables, containing either statistical regularities alone, statistical regularities and speech cues, or no cues. Despite the participants' inability to explicitly detect differences between the speech streams, neural activity differed significantly across conditions, with left-lateralized signal increases in temporal cortices observed only when participants listened to streams containing statistical regularities, particularly the stream containing speech cues. In a second fMRI study, designed to verify that word segmentation had implicitly taken place, participants listened to trisyllabic combinations that occurred with different frequencies in the streams of speech they just heard ("words," 45 times; "partwords," 15 times; "nonwords," once). Reliably greater activity in left inferior and middle frontal gyri was observed when comparing words with partwords and, to a lesser extent, when comparing partwords with nonwords. Activity in these regions, taken to index the implicit detection of word boundaries, was positively correlated with participants' rapid auditory processing skills. These findings provide a neural signature of on-line word segmentation in the mature brain and an initial model with which to study developmental changes in the neural architecture involved in processing speech cues during language learning.

Key words: fMRI; language; speech perception; word segmentation; statistical learning; auditory cortex; inferior frontal gyrus

Introduction

Researchers have long been fascinated by how infants and young children can rapidly "crack" the language code with far greater ease than adults. Although ample behavioral evidence suggests a marked decrease in the ability to acquire a new language with native-like proficiency after puberty (Johnson and Newport, 1989; Weber-Fox and Neville, 2001), little is known about the neural correlates underlying this phenomenon. A continuous process of neural commitment to the statistical and prosodic patterns of the language experienced early in life may account for

the corresponding decrease in the ability to acquire another language later on (Kuhl, 2004). However, the mechanisms by which the brain encodes these statistical and prosodic patterns during the initial stages of language learning have not been examined in adults until now.

Previous functional neuroimaging studies have used several approaches to investigate different aspects of language learning. Some studies examined changes in the pattern of neural activity after a period of intensive training on a novel linguistic task (Friederici et al., 2002; Callan et al., 2003, 2005; Golestani and Zatorre, 2004). Although these studies can address issues of neural plasticity (because they focus on the neural reorganization that results from learning), they do not provide information about the neural systems that subserve the learning process itself. Other investigations explored the neural basis of language learning by examining changes in brain activity that occur while participants learn an artificial grammar in the scanner (Opitz and Friederici, 2003, 2004; Thiel et al., 2003; Hashimoto and Sakai, 2004). Such studies, however, involved language processing in the visual domain and, thus, cannot shed light on the neural mechanisms underlying the learning of the acoustic properties of a language. To investigate such processes, functional magnetic resonance imaging (fMRI) was used to evaluate learning-associated changes in neural activity during initial exposure to

artificial languages presented in the auditory modality, when a listener faces the daunting task of identifying word boundaries in a continuous stream of speech.

Word segmentation is a fundamental aspect of language learning because word boundaries must be identified before any additional linguistic analysis can be performed (Jusczyk, 2002; Saffran and Wilson, 2003). Although speech does not contain acoustic cues that invariably specify word boundaries, research has shown that infants as young as 7.5 months of age can use statistical regularities such as the greater co-occurrence of syllables within words than between words to successfully parse a continuous stream of speech (Saffran et al., 1996a; Aslin et al., 1998). In addition to these transitional probabilities, speech cues available in the input such as stress patterns (i.e., longer duration, increased amplitude, and higher pitch on certain syllables) can also guide word segmentation (Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). By adapting a well established paradigm from the infant behavioral literature, the present research examined the neural correlates of speech parsing. In light of evidence indicating that the ability to process rapid acoustic changes is highly predictive of future linguistic skills (Benasich and Tallal, 2002), the relationship between neural activity indexing successful word segmentation and rapid auditory processing ability was also explored.

Materials and Methods
Participants
Twenty-seven participants (13 female; mean age, 26.63 years; range, 20–44 years) volunteered in this study after giving written informed consent for the experimental protocol approved by the University of California, Los Angeles Institutional Review Board. By report, the participants were right-handed, native English speakers with no history of neurological or psychiatric disorders. All 27 participants completed an fMRI speech stream exposure scan, and 15 of them (six female; mean age, 26.92 years; range, 20–31 years) also underwent an event-related fMRI word discrimination scan immediately after the speech stream exposure scan. All participants completed a behavioral "word discrimination task" outside of the scanner after the scanning session.

Stimuli and tasks
Speech stream exposure task. Participants listened to three counterbalanced streams of nonsense speech supposedly spoken by aliens from three different planets. Participants were not explicitly instructed to perform a task except to listen, given that a recent study has demonstrated that implicit learning can be attenuated by explicit memory processes during sequence learning (Fletcher et al., 2004). As shown in Figure 1, *A* and *B*, the three streams were created by repeatedly concatenating 12 syllables (a different set of 12 syllables was used for each speech stream). In each of the two artificial language conditions, the 12 syllables were used to make four trisyllabic words, following the exact same procedure used in previous infant and adult behavioral studies (Saffran et al., 1996a,b; Aslin et al., 1998; Johnson and Jusczyk, 2001). Each syllable was recorded separately using SoundEdit (Macromedia; Adobe Systems, San Jose, CA), ensuring that the average syllable duration (0.267 s), amplitude (18.2 dB), and pitch (221 Hz) were (1) not significantly different across the experimental conditions and (2) matched to those used previously in the behavioral literature. For each artificial language, the four words were randomly repeated three times to form a block of 12 words, subject to the constraint that no word was repeated twice in a row. Five such different blocks were created, and then this five-block sequence was itself concatenated three times to form a continuous speech stream lasting 2 min and 24 s, during which each word occurred 45 times. For example, the four words "pabiku," "tibudo," "golatu," and "daropi" were combined to form a continuous stream of nonsense speech containing no breaks or pauses (e.g., pabikutibudogolatudaropitibudo. . .). Within the speech stream, transitional probabilities for syllables within a word and across word boundaries were 1 and 0.33, respectively. Thus, as the
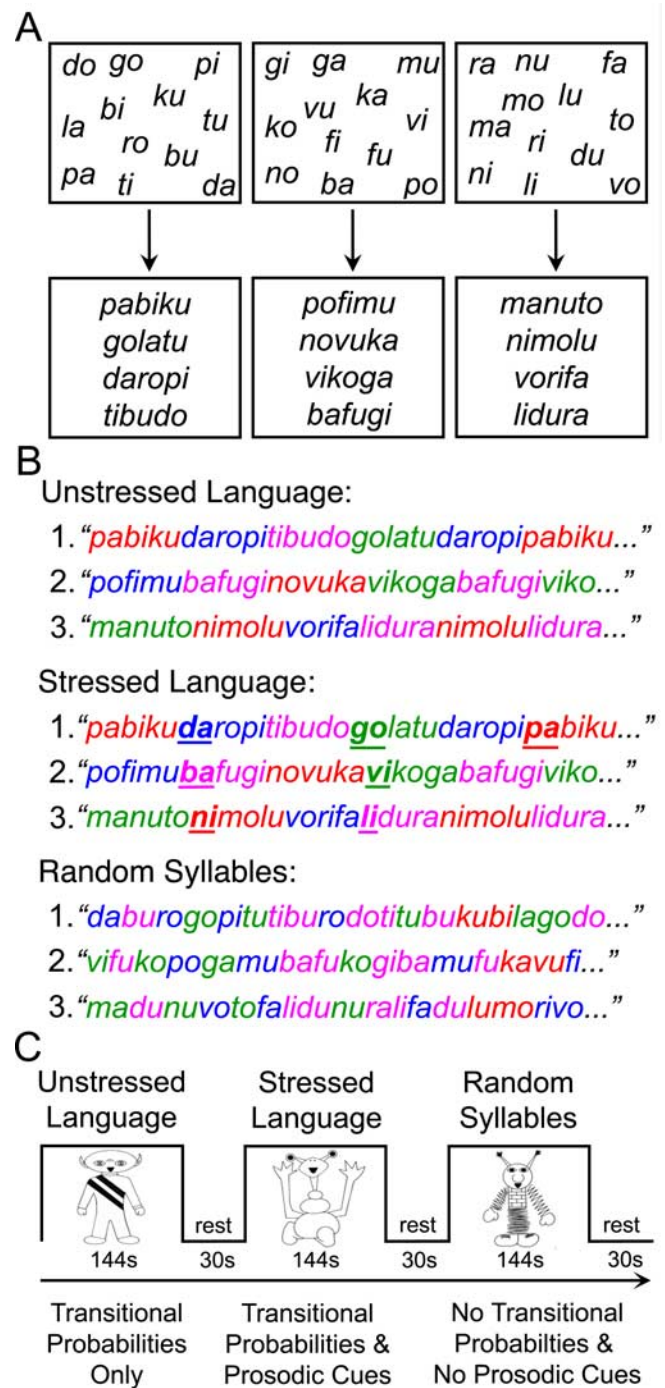
Figure 1. In the speech stream exposure task, three sets of 12 syllables were used to create three sets of four words (*A*). The unstressed and stressed languages were formed by concatenating these words to form two artificial languages, whereas the random syllables stream was formed by pseudorandomly concatenating individual syllables (*B*). Participants listened to three counterbalanced speech streams (*C*) containing statistical regularities (unstressed language), statistical regularities and prosodic cues (stressed language), and no cues to guide word segmentation (random syllables).

words were repeated, transitional probabilities could be computed and used to segment the speech stream. In the "unstressed language condition" (U), the stream contained only transitional probabilities as cues to word boundaries. In the "stressed language condition" (S), the speech stream contained transitional probabilities, as well as speech, or prosodic, cues introduced by adding stress to the initial syllable of each word, one-third of the time it occurred. Stress was added by slightly increasing

the duration (0.273 s), amplitude (16.9 dB), and pitch (234 Hz) of these stressed syllables. At the same time, these small increases were offset by minor reductions in these parameters for the remaining syllables within the stressed language condition, which ensured that the mean duration, amplitude, and pitch would not be reliably different across the three experimental conditions. The initial syllable was stressed because 90% of words in conversational English have stress on their initial syllable (Cutler and Carter, 1987).

A "random syllables condition" (R) was also created so as to identify activity associated with the actual process of word segmentation, as afforded by the presence of statistical and prosodic cues available in the two artificial language conditions (U + S), as opposed to activity related to merely listening to a series of concatenated syllables. In this condition, the 12 syllables were not arranged into four words as in the two artificial language conditions; rather, these syllables were arranged pseudorandomly such that no three-syllable string was repeated more than twice in the stream (the frequency with which two-syllable strings occurred was also minimized). Therefore, in this condition, neither statistical regularities nor prosodic cues were afforded to the listener to aid speech parsing. Thus, as depicted in Figure 1C, each participant listened to three 144 s speech streams (R, U, and S) interspersed between 30 s of resting baseline. Short samples of these speech streams are available in supplemental audio clips 1–3 (available at www.jneurosci.org as supplemental material). The order of presentation of the three experimental conditions was counterbalanced across participants according to a "Latin square" design.

It is important to note the following: (1) the same number of syllables was repeated the same number of times across all three conditions, although it was only in the two artificial language conditions that cues were available to guide word segmentation; (2) across participants, each set of 12 syllables was used with the same frequency in each condition, thus ensuring that any difference between conditions would not be attributable to different degrees of familiarity with a given set of syllables; (3) to guard against the possibility that the computation of the transitional probabilities between the syllables chosen to form the words in the two artificial languages might be influenced by previous experience with the transitional probabilities between these syllables in English, three different versions of each language were created by rotating the position of the syllables within the words of each language (e.g., "pabiku" in one version became "kubipa" and "bipaku" in the others, with each word being used equally often across participants); and (4) although the length of the activation blocks used in this task was unconventional for an imaging study and could have resulted in decreased power to detect reliable differences between conditions, we opted to adhere to the paradigm used in previous behavioral studies in light of preliminary analyses conducted on the first 12 subjects demonstrating that sufficient power was nevertheless achieved. Furthermore, effect sizes, computed to address the robustness of the analyses performed, demonstrated that rather large effects were obtained (Cohen's $d > 0.80$).

Word discrimination task. Fifteen participants also completed a second fMRI task to provide an implicit index of successful word segmentation in lieu of the assessment procedure (i.e., head-turn novelty preference) used in the infant behavioral studies. In this mixed block/event-related fMRI design, participants were presented with trisyllabic combinations from the "speech stream exposure task" and were simply told to listen to what might have been words in the artificial languages they just heard. Based on evidence from previous behavioral studies, participants were not expected to be able to explicitly identify whether these trisyllabic combinations were words in the artificial languages after such a short exposure to the streams (Saffran et al., 1996b; Sanders et al., 2002). Accordingly, participants were not asked to make an explicit judgment in the scanner so as not to have explicit task demands confound the activity associated with the implicit processing of the stimuli. As shown in Figure 2, A and B, trisyllabic combinations were presented in three activation blocks, with each block using the set of syllables that corresponded to those used in one of the three speech streams heard previously by a participant. Within each block, participants listened to the four "words" (W) used to create the artificial language streams (e.g., "golatu", "daropi") as well as to four "partwords" (PW), that is, trisyllabic combina-
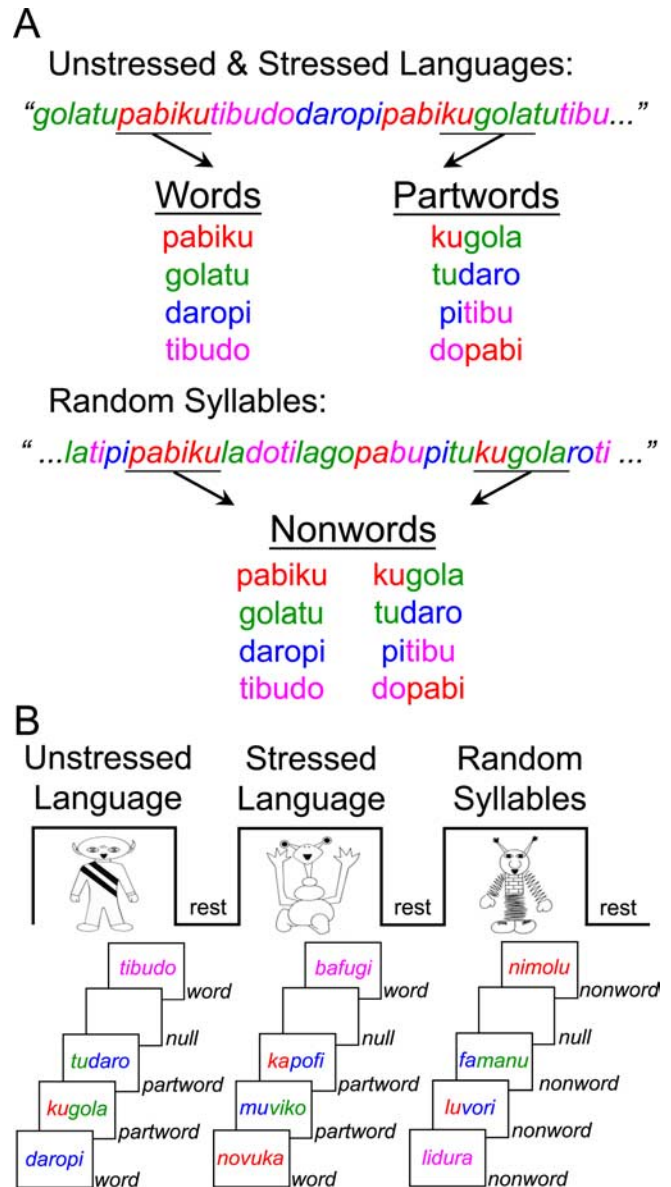

Figure 2. In the word discrimination task, participants listened to words and partwords from the unstressed and stressed language streams and to nonwords from the random syllables stream (*A*). These trisyllabic combinations were presented in three activation blocks, with each block using the set of syllables that corresponded to those used in one of the three speech streams heard previously by a participant (*B*).

tions formed by grouping syllables from adjacent words within the speech streams (e.g., the partword "tudaro" consisted of the last syllable of the word "golatu" and the first two of the adjacent word "daropi" within the stream ". . . gola**tudaro**pi. . . "). As in the previous behavioral studies, the transitional probabilities between the first and second syllables in the words were appreciably higher than those between the first two syllables in the partwords (1 as opposed to 0.33) as a result of the words and partwords having occurred within the speech stream 45 and 15 times, respectively. Each word and partword was repeated three times and interspersed with null events. The same pseudorandomized order of presentation for words and partwords was used in each block such that no word or partword was repeated twice in a row and no more than three words, partwords, or null events occurred consecutively. The words and partwords from the stressed language condition were presented in their unstressed version, such that there were no differences in duration, amplitude, or pitch between any of the syllables used in this task. The trisyllabic combinations used in the block corresponding to the random syl-

lables condition effectively served as "nonwords" (NW), because these combinations had occurred only one time during the random syllables stream, in which no statistical or prosodic cues were afforded to the listener (transitional probability between the first two syllables equal to 0.1).

Behavioral tasks. To investigate whether participants were able to explicitly discriminate between words and partwords, behavioral measures (response times and accuracy scores) for the word discrimination task were collected outside of the scanner. Each participant listened to the word stimuli and responded yes or no as to whether they thought each trisyllabic combination could be a word in the artificial languages they heard previously. All 27 participants completed this behavioral testing. All but one participant also completed a computerized version of the Tallal Repetition Test, a hierarchical series of subtests shown to assess several aspects of central auditory processing, including rapid auditory processing skills (Tallal and Piercy, 1973). This test uses a two-alternative forced-choice design in which participants are operantly trained to respond to two complex tones by clicking on two different squares on the computer screen. After participants learn the tone–square association, they are presented with progressively longer sequences of complex tones and must subsequently indicate the temporal order of such sequences. The accuracy scores from the SerialMemory3Fast (SM3Fast) subscale of this test were used in this study because maximum variability is observed among participants at this difficulty level. In this subscale, the stimuli were two 75 ms complex tones, with fundamental frequencies of 100 and 300 Hz. Random combinations of these two tones were presented as tone triplets, consisting of three elements (e.g., tone 1-2-1, 2-1-1, etc.) separated by varying interstimulus intervals of silence (0, 10, 60, or 150 ms). Participants listened to 24 such three-element trials, indicating after each trial the order of the stimulus pattern by clicking on the squares in the same order as the tones were presented.

fMRI data acquisition

Functional images were collected using a Siemens Allegra 3 Tesla head-only MRI scanner. A two-dimensional spin-echo scout [repetition time (TR), 4000 ms; echo time (TE), 40 ms; matrix size, 256 × 256; 4 mm thick; 1 mm gap] was acquired in the sagittal plane to allow prescription of the slices to be obtained in the remaining scans. For each participant, a high-resolution structural T2-weighted echo-planar imaging (EPI) volume [TR, 5000 ms; TE, 33 ms; matrix size, 128 × 128; field of view (FOV), 20 cm; 36 slices; 1.56 mm in-plane resolution; 3 mm thick] was acquired coplanar with the functional scans to allow for spatial registration of each participant's data into a standard coordinate system. For the speech stream exposure task, one functional scan lasting 8 min and 48 s was acquired covering the whole cerebral volume (174 images; EPI gradient echo sequence; TR, 3000 ms; TE, 25 ms; flip angle, 90°; matrix size, 64 × 64; FOV, 20 cm; 36 slices; 3.125 mm in-plane resolution; 3 mm thick; 1 mm gap). For the word discrimination task, a second functional scan lasting 3 min and 30 s was acquired that also covered the whole brain (103 images; EPI gradient echo sequence; TR, 2000 ms; TE, 25 ms; flip angle, 90°; matrix size, 64 × 64; FOV, 20 cm; 36 slices; 3.125 mm in-plane resolution; 3 mm thick; 1 mm gap).

Participants listened to the auditory stimuli through a set of magnet-compatible stereo headphones (Resonance Technology, Northridge, CA). Stimuli were presented using MacStim 3.2 psychological experimentation software (CogState, West Melbourne, Victoria, Australia).

fMRI data analysis

Using automated image registration (Woods et al., 1998a,b), functional images for each participant were (1) realigned to each other to correct for head motion during scanning and coregistered to their respective high-resolution structural images using a six-parameter rigid body transformation model and a least-square cost function with intensity scaling, (2) spatially normalized into a standard stereotactic space (Woods et al., 1999) using polynomial nonlinear warping, and (3) smoothed with a 6 mm full-width at half-maximum isotropic Gaussian kernel to increase the signal-to-noise ratio. Statistical analyses were implemented in SPM99 (Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). For each participant, contrasts of interest

were estimated according to the general linear model using a canonical hemodynamic response function. For the speech stream exposure task, the exponential decay function in SPM99 (which closely approximates a linear function) was also used to model changes that occurred within each activation block as a function of exposure to the speech stream. Contrast images from these fixed effects analyses were then entered into second-level analyses using random effects models to allow for inferences to be made at the population level (Friston et al., 1999). All reported activity survived correction for multiple comparisons at the cluster level ($p < 0.05$, corrected) and $t > 3.3$ for magnitude ($p < 0.001$, uncorrected), unless otherwise noted. Small-volume correction at the cluster level with a sphere of 1 cm radius around the peak of activation was used in the putamen for the speech stream exposure task because of an a priori hypothesis of basal ganglia involvement during sequence learning.

For the speech stream exposure task, separate one-sample t tests were implemented for each condition (unstressed language, stressed language, and random syllables vs resting baseline) to identify blood oxygenation level-dependent (BOLD) signal increases associated with listening to each speech stream. Direct comparisons between each condition, as well as between the two artificial language conditions (U + S) and the random syllables condition, were also implemented to examine differential activity related to the presence of statistical and prosodic cues within the artificial language streams. Specifically, in an effect akin to repetition priming, the detection of word boundaries (afforded by the statistical regularities and prosodic cues available in the artificial languages only) was expected to yield overall less activity for the artificial language streams within language-relevant frontotemporal regions and the basal ganglia, which are known to be involved in sequence learning (Poldrack et al., 2005).

In addition, because word segmentation is thought to take place as a function of exposure to the language streams, signal increases over time were expected in left-hemisphere language cortices while listening to the artificial languages only, with a larger effect predicted for the stressed language condition in which both statistical and prosodic cues were available to guide word segmentation. A region-of-interest (ROI) analysis was then conducted in the temporal cortices in which such effects were observed to examine whether these signal increases over time for the artificial languages were indeed left-lateralized. This functionally defined ROI included all voxels showing reliable signal increases at the group level in either the left hemisphere (LH) or right hemisphere (RH), as well as their respective counterparts in the opposite hemisphere. For each participant, a laterality index was computed based on the number of voxels showing increased activity over time within these symmetrical ROIs in the LH and RH (number of voxels activated in the LH - number of voxels activated in the RH/number of voxels activated in the LH + RH).

Because the ability to identify words that had occurred within the speech streams should be related to activity observed during exposure to the artificial languages if this activity truly reflects word segmentation, regression analyses were also conducted to assess the relationship between accuracy scores on the behavioral post-scanner word discrimination test and neural activity associated with listening to the two artificial languages compared with the random syllables during the speech stream exposure task.

For the word discrimination task, separate one-sample t tests were implemented to identify changes in the BOLD signal associated with listening to the words, partwords, and nonwords that had occurred with different frequencies during the speech stream exposure task. Direct comparisons between each condition were then implemented to test the hypothesis that differential activity in language networks would be observed for words, partwords, and nonwords, as would be expected if word segmentation had indeed occurred during the speech stream exposure task. Specifically, this effect could manifest as stronger signal increases in response to words (than partwords and nonwords) if these were recognized as having "word-like" status, or, alternatively, the opposite pattern could be observed because of a novelty effect for nonwords. The SPM99 toolbox MarsBaR (http://marsbar.sourceforge.net/) was used to extract parameter estimates for each participant from regions

Table 1. Behavioral performance on the word discrimination task

	Accuracy (% correct)		Response time (s)	
	Words	Partwords	Words	Partwords
Unstressed language	58.3 (21.2)	39.8 (22.2)	1.531 (0.241)	1.574 (0.286)
Stressed language	47.2 (25.3)	46.3 (25.7)	1.526 (0.295)	1.539 (0.257)
Random syllables	55.6 (26.3)	45.4 (26.9)	1.512 (0.260)	1.568 (0.335)

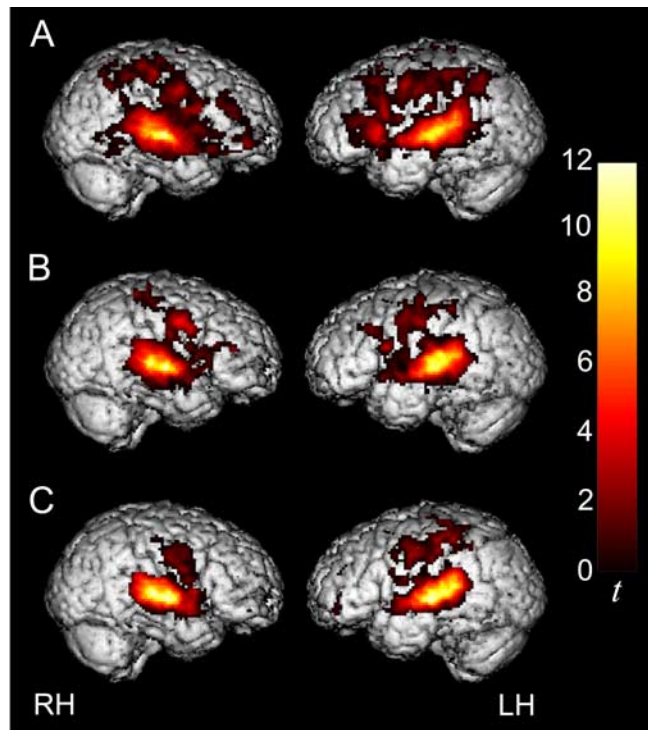Values presented as mean (SD).

Figure 3. Extensive activity in temporal, frontal, and parietal cortices, as well as in the putamen, was observed when participants listened to the random syllables stream (**A**), the unstressed language stream (**B**), and the stressed language stream (**C**) compared with resting baseline, with a more focal pattern of activity observed when more cues were available in the speech stream. Activation maps are displayed at a threshold of $t > 1.71$, with correction for multiple comparisons at the cluster level ($p < 0.05$, corrected).

that were significantly active in response to words, partwords, and nonwords.

Finally, given that rapid auditory processing skills are highly predictive of future language outcomes (Benasich and Tallal, 2002), a regression analysis was conducted to examine the relationship between participants' rapid auditory processing skills and the ability to implicitly segment the artificial language streams, as indexed by activity observed in response to words in the word discrimination task. Based on previous findings linking the left middle frontal gyrus (MFG) to the processing of rapid acoustic changes (Temple, 2002), a positive correlation was expected between accuracy scores on the SM3Fast subscale of the Tallal Repetition Test and neural activity associated with listening to words in the word discrimination task.

Results
Behavioral results
Accuracy and reaction times on the behavioral word recognition test conducted immediately after the fMRI scans are reported in Table 1. Not surprisingly, given the findings of previous behavioral research (Saffran et al., 1996b; Sanders et al., 2002), participants were unable to explicitly recognize which trisyllabic combinations might have been words in the artificial languages they heard during the speech stream exposure task, as demonstrated by accuracy scores not different from chance for either the stressed or unstressed language conditions. Similarly, there were no significant differences in reaction time to items from within the unstressed language, stressed language, or random syllables streams.

Speech stream exposure task
Random effects models were used to identify changes in the BOLD signal associated with listening to the unstressed language, stressed language, and random syllables during the speech stream exposure task. Compared with resting baseline, listening to all three streams of continuous speech activated a large-scale bilateral neural network including canonical language areas in the LH and their RH counterparts, with extensive activation in temporal, frontal, and parietal cortices as well as in the putamen (Fig. 3; Table 2). Overall, a pattern of more focal activity (i.e., less activation in terms of spatial extent) was observed when participants listened to the artificial languages (U + S) than when they listened to the stream of random syllables (Table 2). Similarly, more focal activity was observed when participants listened to the stressed language than to the unstressed language (Table 3), indicating that the presence of statistical cues in both artificial language conditions, as well as the addition of prosodic cues in the stressed language, afforded more efficient neural processing.

Because transitional probabilities and the related identification of word boundaries are computed on-line during the course of the activation conditions, regions in which activity might be increasing as a function of exposure to the speech streams were examined. A statistical contrast modeling increases in activity over time for the artificial languages (U ↑ + S ↑) revealed significant bilateral signal increases in superior temporal gyrus (STG) and transverse temporal gyrus, extending into the supramarginal gyrus (SMG) in the LH (Fig. 4A,B; Table 4). A region-of-interest analysis indicated that this increasing activity during exposure to the unstressed and stressed language conditions was reliably stronger in the LH than in the RH, as indicated by a significantly larger number of voxels showing this effect in the LH than in the RH (LH, 235 vs RH, 180; $F_{(1,26)} = 4.622$; $p < 0.05$). In contrast, no reliable signal increases over time were observed during exposure to the random syllables condition, as can be seen in the time series shown in Figure 4B. A contrast directly comparing signal increases occurring during exposure to the two artificial languages versus signal increases occurring during exposure to the random syllables condition [(U ↑ + S ↑) − R ↑] confirmed that activity in bilateral middle temporal gyrus (MTG) and STG, and left SMG showed significantly greater increases as a function of exposure to the two artificial language conditions than the random syllables condition (Table 4). This pattern held when comparing signal increases during each artificial language condition versus the random syllables condition (U ↑ − R ↑ and S ↑ − R ↑) (Fig. 5A,B; Table 5), with stronger signal increases for the comparison of the stressed language versus random syllables than for the unstressed language versus random syllables [(S ↑ − R ↑) − (U ↑ − R ↑)] in left STG and the right temporoparietal junction (Fig. 5C; Table 5).

A statistical contrast modeling decreases in activity over time for the artificial language conditions (U ↓ + S ↓) revealed activity in ventromedial prefrontal cortex (Talairach coordinates 4, 56, 8; $t = 5.26$, corrected for multiple comparisons at the cluster level). Interestingly, at more liberal thresholds ($t > 1.71$; $p < 0.05$,

uncorrected; minimum cluster size, 5 voxels), signal decreases were also observed in several other regions (i.e., left MTG, bilateral inferior parietal lobule (IPL), right middle and medial frontal gyri, anterior cingulate, and left putamen) in which overall greater activity was observed for the random syllables versus artificial language conditions [(R − (S + U)]. Conversely, at these more liberal thresholds, greater activity was observed for the artificial languages versus random syllables condition [(S + U) − R)] in the left STG and IPL, that is, within those regions in which significant increases over time were observed for the artificial languages. For the random syllables condition (R ↓), signal decreases over time were only observed in the left insula and IPL (Talairach coordinates −34, −24, −2, $t = 5.3$ and −50, −36, 50, $t = 3.75$, respectively, both corrected for multiple comparisons at the cluster level).

Correlation with behavioral word discrimination accuracy

Participants' accuracy at discriminating words on the post-scanner behavioral word discrimination task was positively correlated with activity in the left STG as participants listened to the stressed and unstressed languages compared with random syllables during the speech stream exposure task (Fig. 6; Table 6). Importantly, this effect was observed in the same region showing increasing activity over time while listening to the artificial languages.

Word discrimination task

Random effects models were used to identify changes in the BOLD signal associated with listening to words, partwords, and nonwords presented during the three blocks of the word discrimination task. Activity in response to hearing both words and partwords (which had occurred 45 and 15 times, respectively, during the speech stream exposure task) was summed across the unstressed and stressed language blocks. (In keeping with the head-turn novelty preference paradigm used in the developmental literature on word segmentation, only 12 words and 12 partwords were presented within each language block. Collapsing across the language blocks yielded 24 events for words and 24 for partwords to be contrasted with the 24 events for the nonwords from the random syllable block.) Activity for the nonwords corresponded to trisyllabic combinations presented during the random syllables block (these combinations had occurred only once during the random syllables condition of the speech stream exposure task). A statistical contrast comparing BOLD signal associated with listening to words versus

Table 2. Peaks of activation during exposure to the speech streams

Anatomical regions	BA		All activation > rest				Random syllables > artificial languages			
			x	y	z	t	x	y	z	t
Transverse temporal gyrus	41	L	−46	−32	12	9.92				
	41	R	48	−24	8	11.45				
Superior temporal gyrus	22	L	−66	−30	12	10.52	−50	−16	2	7.28
	22	R	48	−12	4	11.72	40	−8	−4	3.79
	42	L	−56	−26	8	12.47				
	42	R	56	−14	6	12.05				
Middle temporal gyrus	21	L	−68	−28	2	8.18	−40	−14	−12	4.09
	21	R	58	−14	−2	10.65	52	−8	−12	3.96
	22	L	−68	−34	4	8.80				
	37	R					54	−52	0	3.45
Inferior parietal lobule	40	L	−52	−38	48	3.88	−54	−48	32	4.95
	40	R					52	−54	38	4.18
Inferior frontal gyrus	44	L	−40	48	0	4.11	−44	16	10	6.76
	44	R	44	10	10	4.62				
	45	L	−46	24	2	3.79	−52	24	8	3.73
	45	R	48	26	4	5.41				
	47	L	−42	42	8	3.74				
	47	R	38	20	−8	4.14				
Middle frontal gyrus	6	L	−6	4	60	6.27				
	9	L	−48	20	28	3.80				
	10	L	−28	48	−2	3.53				
	10	R	32	46	−6	3.91	34	46	−2	3.46
	46	R					40	34	26	4.08
Precentral gyrus	4	R	30	−28	48	4.20	24	−18	60	3.94
	6	L	−50	−2	44	6.35				
	6	R	48	−2	38	6.78				
Anterior cingulate gyrus	32	L	−8	44	14	4.67	−8	12	34	3.74
	32	R	14	46	−2	3.38	4	12	34	3.30
Medial frontal gyrus	6	R	2	−26	64	3.64	8	−30	52	4.63
	8	L	−4	34	42	3.68				
	9	L	−8	52	14	4.21				
	9	R					4	38	26	4.08
	10	L	−10	48	2	5.19				
	10	R	4	58	4	4.17				
Superior frontal gyrus	10	L	−20	56	8	3.76				
	10	R	18	56	8	3.43				
Insula		L					−42	16	0	4.82
		R					40	−8	−4	3.79
Putamen		L	−28	−4	6	4.08	−28	4	14	3.87
		R	20	−4	4	3.37				
Thalamus		R					6	−12	2	4.38
Cerebellum		L					−30	−62	−28	3.92

Activity thresholded at $t > 3.30$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$), except for the putamen in which a small-volume correction was applied. L, Left; R, right.

Table 3. Peaks in which activity was greater during exposure to the unstressed language stream than the stressed language stream

Anatomical regions	BA		Unstressed language > stressed language			
			x	y	z	t
Inferior frontal gyrus	45	R	34	28	2	3.69
	47	R	32	46	−4	3.53
Putamen		L	−20	4	4	4.28
		R	20	4	4	3.14*

Activity thresholded at $t > 3.30$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$). * indicates only survives, $p < 0.002$. R, Right; L, left.

nonwords (W − NW) revealed significantly greater activity for words in left inferior frontal gyrus (IFG) and MFG (Fig. 7A; Table 7). Reliably greater activity, albeit to a lesser extent, was also observed in these same regions when comparing words versus partwords (W − PW) (Fig. 7B; Table 7) and partwords versus nonwords (PW − NW) (Table 7). The partwords versus nonwords comparison also revealed significantly greater activity in the superior parietal lobule. Importantly, no reliably greater cor-
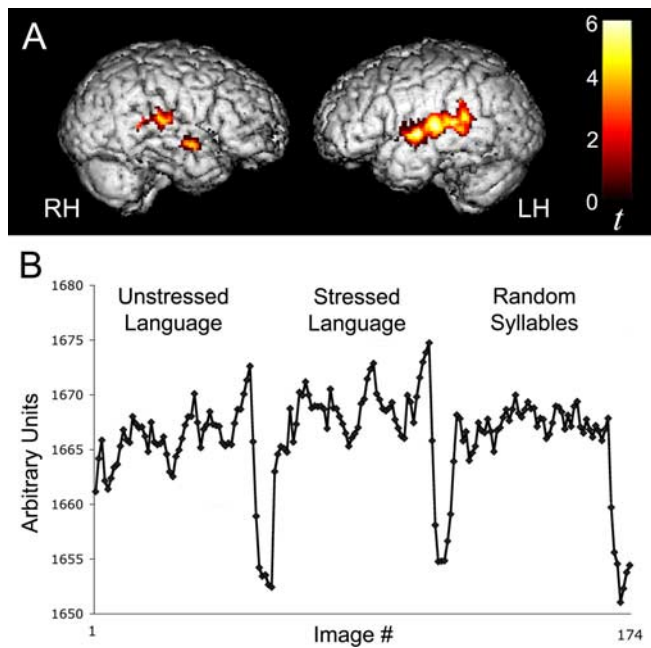
Figure 4. Activity within temporal cortices was found to increase over the course of listening to the speech streams for both artificial language conditions (U ↑ + S ↑), thresholded at $t >$ 1.71, with correction for multiple comparisons at the cluster level ($p < 0.05$, corrected) (**A**). The time series showing signal increases in temporal cortices during exposure to the three speech streams illustrates the significant increases observed for the unstressed and stressed language conditions but not for the random syllables condition (**B**).
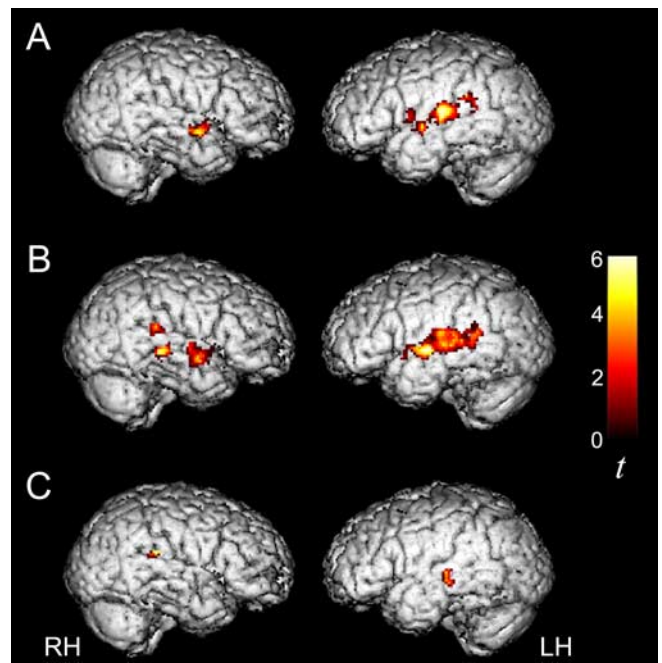

Figure 5. Activity within temporal cortices was found to increase during exposure to the unstressed language stream (**A**) and the stressed language stream (**B**) to a greater extent than during exposure to the random syllables stream (U ↑ − R ↑ and S ↑ − R ↑). Activity within the left STG and right temporoparietal junction was found to increase during exposure to the stressed language stream to a greater extent than during exposure to the unstressed language stream (S ↑ − U ↑) (**C**). Activation maps are displayed at a threshold of $t >$ 1.71, with correction for multiple comparisons at the cluster level ($p < 0.05$, corrected).

tical activity was observed for any of the reverse statistical comparisons (NW − W, NW − PW, and PW − W). Despite the limited number of words and partwords within each language block, when the above analyses were conducted for the stressed and unstressed language blocks separately, a virtually identical pattern of results was observed (albeit with smaller cluster size).

Additional analyses were conducted to determine whether differential activity would be observed as participants listened to words that had occurred in the speech stream when both prosodic and statistical cues were available to guide word segmentation (stressed language) as opposed to words that had occurred in the speech stream containing only statistical cues (unstressed language). Comparing activity observed during the stressed language block with that observed during the unstressed language block [S(W + PW) − U(W + PW)] revealed that left MFG and the right IPL (Talairach coordinates −48, 12, 34, $t = 2.85$ and 36, −50, 38, $t = 3.43$, respectively, corrected for multiple comparisons at the cluster level) showed greater activation when partici-

pants listened to words and partwords from within the speech stream that had contained additional prosodic cues (stressed language) than to words and partwords from within the stream containing statistical regularities alone (unstressed language). No reliable activity was detected for the reverse comparison [U(W + PW) − S(W + PW)].

Correlation with rapid auditory processing skills

Participants' performance on the Tallal Repetition Test was positively correlated with left MFG activity while participants listened to words during the word discrimination task (Fig. 8). That is, the better the participants' rapid auditory processing skills, as indexed by accuracy on the SM3Fast subscale of the Tallal Repetition Test (mean ± SD, 80.3 ± 19.6%), the greater the activity observed in left MFG during the word discrimination task when listening to trisyllabic combinations that had occurred 45 times

Table 4. Peaks of activation for regions in which activity increased as a function of exposure to both artificial language speech streams (stressed and unstressed languages combined)

Anatomical regions	BA		Increases in artificial languages (U ↑ + S ↑)				Increases in artificial languages > random syllables (U ↑ + S ↑ > R ↑)			
			x	y	z	t	x	y	z	t
Superior temporal gyrus	22	L	−62	−8	4	4.42	−46	−44	18	4.35
	22	R	46	−2	−4	4.91	52	−2	−2	3.74
	42	L	−50	−30	10	5.99	−64	−28	12	4.25
	42	R	44	−30	14	3.51				
Transverse temporal gyrus	41	L	−44	−26	12	4.75				
	41	R	30	−28	14	3.35				
Inferior parietal lobule	40	L	−60	−42	28	3.70	−60	−42	28	3.59

Activity thresholded at $t > 3.30$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$). L, Left; R, right.

Table 5. Peaks of activation for regions in which activity increased as a function of exposure to the unstressed and stressed language speech streams

Anatomical regions	BA		Increases in unstressed language > random syllables				Increases in stressed language > random syllables				Increases in stressed language > unstressed language			
			(U ↑ > R ↑)				(S ↑ > R ↑)				(S ↑ − R ↑ > U ↑ − R ↑)			
			x	y	z	T	x	y	z	t	x	y	Z	t
Superior temporal gyrus	22	L	−46	−44	18	3.41	−42	−42	12	3.65	−62	−24	6	2.82*
	22	R	48	0	−4	3.99	44	−2	−6	3.57	44	−38	20	3.90
	42	L					−66	−28	8	4.19				
	42	R												
Transverse temporal gyrus	41	L					44	−30	16	3.52				
	41	R												
Inferior parietal lobule	40	L	−66	−22	16	3.70	−62	−20	16	3.70				

Activity thresholded at (t > 3.30 (p < 0.001), corrected for multiple comparisons at the cluster level (p < 0.05). * indicates only survives, p < 0.004. L, Left; R, right.
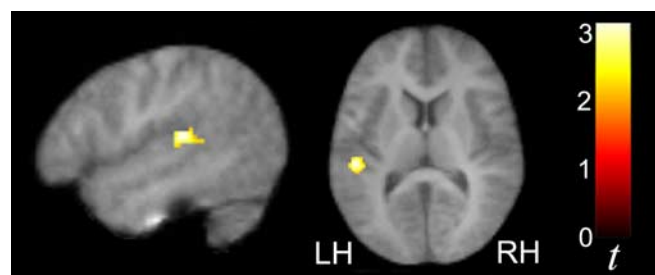

Figure 6. Regression analysis revealed that left superior temporal gyrus activity associated with listening to the artificial language conditions (compared with the random syllables condition) during the speech stream exposure task was significantly correlated with participants' ability to discriminate words that occurred in the artificial languages, as indexed by their accuracy on the behavioral post-scanner word discrimination task [peak at Talairach coordinates −50, −30, 14, putative Brodmann area 42; t = 2.94, with correction for multiple comparisons at the cluster level (p < 0.05, corrected)].

during the stressed and unstressed language conditions of the speech stream exposure task.

Discussion

The present research investigated the neural correlates of word segmentation, a crucial process in the early stages of language learning. Over the past decade, milestone studies in the infant behavioral literature have revolutionized our thinking about the mechanisms underlying language acquisition, showing that the brain possesses striking computational abilities that allow young infants to rapidly identify word boundaries by extracting statistical and prosodic patterns within continuous speech (Saffran et al., 1996a; Aslin et al., 1998; Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). The neural architecture underlying this computational feat, however, has not yet been identified. By adapting the paradigm used in the infant literature, this study explored how the brain processes the statistical and prosodic speech cues available in the input to crack the code of a new artificial language.

Despite the somewhat unusual fMRI design, the frontotemporal networks activated while participants listened to streams of concatenated syllables are consistent with those observed in previous studies involving the presentation of similar speech-like stimuli and spectrally complex acoustic information (Binder et al., 2000; Hickok and Poeppel, 2000; Johnsrude et al., 2000; Scott et al.; 2000; Vouloumanos et al., 2001; Gandour et al., 2002; Jäncke et al., 2002; Zatorre et al., 2002; Gelfand and Bookheimer, 2003; Scott and Johnsrude, 2003; Dehaene-Lambertz et al., 2005) (Fig. 3; Table 2). The observed activity in motor speech areas is also in line with a recent study showing considerable overlap between the areas active during passive listening to syllables and those active during overt syllable production (Wilson et al., 2004). Moreover, our finding of basal ganglia activity is in agreement with previous studies involving sequence learning across different modalities (Poldrack et al., 2005; Doyon et al., 2003; Saint-Cyr, 2003; Van der Graaf et al., 2004).

The more focal pattern of activity observed while participants listened to the artificial languages, as opposed to the stream of random syllables, suggests that the statistical and prosodic cues to word segmentation present only in the artificial languages facilitated the parsing of the speech streams, thus affording more efficient neural processing (Table 2). This interpretation is line with other studies reporting less activity for easier or better-learned tasks (Petersen et al., 1998; Landau et al., 2004; Casey et al., 2005; Kelly and Garavan, 2005). The overall greater activity associated with listening to the stream of random syllables, in which the order of occurrence of syllables could not be predicted, is also consistent with a study showing greater activity in superior temporal cortices in response to sequences of visual stimuli with low temporal predictability versus those with high predictability (Bischoff-Grethe et al., 2000).

During exposure to three continuous streams of concatenated syllables, left-lateralized BOLD signal increases in temporal and inferior parietal cortices occurred only when participants listened to the two streams containing statistical regularities, particularly the stream also containing prosodic cues (Figs. 4, 5; Tables 4, 5). Because the artificial languages and random syllables stream differed only in the presence of cues to word boundaries, the signal increases observed for both artificial languages along the superior

Table 6. Peaks of activation for regions in which activity during listening to the speech streams was related to word discrimination accuracy

Anatomical regions	BA		Artificial languages > random syllables				Stressed language > random syllables				Unstressed language > random syllables			
			(U + S > R)				(S > R)				(U > R)			
			x	y	z	t	X	y	z	t	x	y	Z	t
Superior temporal gyrus	42	L	−50	−30	14	2.94	−46	−32	12	3.61	−52	−28	12	2.38
Middle temporal gyrus	21	L					−68	−44	4	3.37				

Activity thresholded at t > 2.3 (p < 0.01), corrected for multiple comparisons at the cluster level (p < 0.05). L, Left.
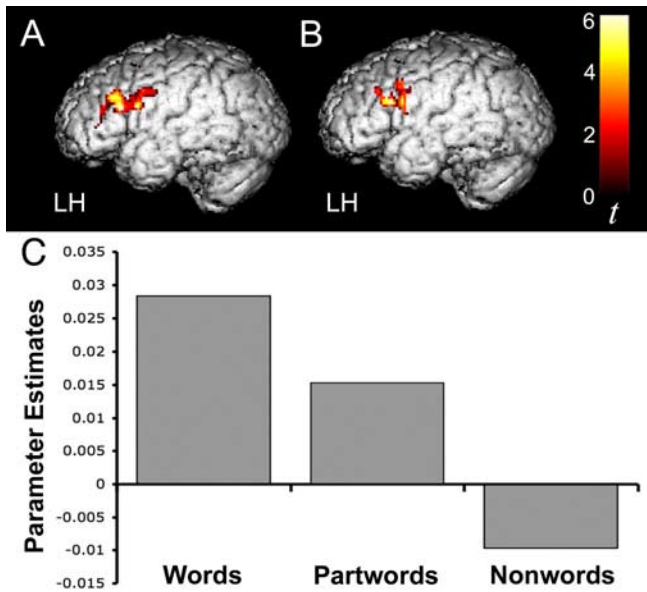
Figure 7. Reliable activity was observed within the posterior left inferior and middle frontal gyri as participants listened to words compared with nonwords (W − NW) (***A***) and partwords (W − PW), thresholded at *t* > 2.48, with correction for multiple comparisons at the cluster level (*p* < 0.05, corrected). The bar graph shows the parameter estimates of activity within these regions while participants listened to words, partwords, and nonwords (***C***).
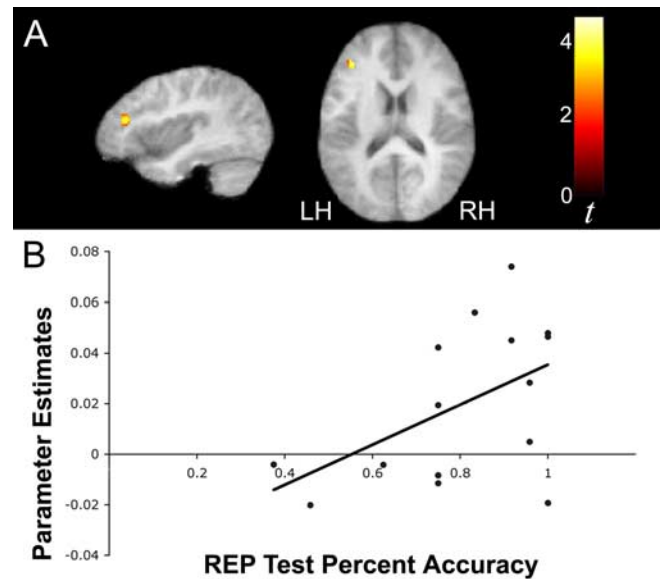

Figure 8. Regression analysis revealed that middle frontal gyrus activity associated with listening to words was significantly correlated with participants' rapid auditory processing skills, as measured by accuracy on the Tallal Repetition (REP) Test [peak at Talairach coordinates −40, 36, 16, putative Brodmann area 46; *t* = 4.37, with correction for multiple comparisons at the cluster level (*p* < 0.05, corrected)] (***A***). The scatter plot illustrates the positive correlation between accuracy on the REP Test and activity within the left middle frontal gyrus (*r* = 0.504; *p* < 0.028) (***B***).

temporal gyri may reflect the ongoing computation of transitional probabilities (Blakemore et al., 1998; Bischoff-Grethe et al., 2000), whereas the increases in activity in the left supramarginal gyrus may reflect the development of phonological representations for the "words" in the artificial languages (Gelfand and Bookheimer, 2003; Xiao et al., 2005). The stronger signal increases observed in these regions, as well as in primary auditory cortices, for the stressed language compared with the unstressed language, suggest that the presence of prosodic cues did aid speech parsing, in line with previous behavioral data in both infants and adults (Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003) (Table 5). Although several studies have reported decreases in activity as participants learn to perform a linguistic task (Thiel et al., 2003; Golestani and Zatorre, 2004; Noppeney and Price, 2004) and, indeed, some decreases were also observed in the present study, our findings are consistent with previous reports of increased activity as a function of learning in primary and secondary sensory and motor areas, as well as in regions involved in the storage of task-related cortical representations (for review, see Kelly and Garavan, 2005).

Our finding of signal increases in temporal and inferior parietal cortices during exposure to the artificial languages indicates that initial word segmentation had occurred, despite the fact that, overall, participants could not reliably identify the words used to create the speech streams in a post-scan behavioral test. This interpretation is strongly supported by the positive relationship

found between activity within the left STG during exposure to the artificial languages (compared with the random syllables stream) and participants' ability to discriminate words in the post-scanner behavioral word discrimination task (Fig. 6; Table 6). Additionally, although neuroimaging studies of learning have typically examined changes in neural activity associated with observable behavioral changes (indexed by explicit or implicit measures), several other studies have reported changes in neural activity that occur in the absence of any overt change in task performance (Shadmehr and Holcomb, 1997; Jaeggi et al., 2003; Landau et al., 2004; Kelly and Garavan, 2005). Neural changes that precede behavioral changes have also been shown in event-related potential (ERP) studies across a variety of linguistic tasks (Shestakova et al., 2003; McLaughlin et al., 2004). Notably, an ERP study of word segmentation has also demonstrated that, although adults required prolonged exposure and training to explicitly identify words within a speech stream, they displayed an increased N100 component to word initial syllables (taken to index the implicit detection of word boundaries) long before showing evidence of explicit knowledge (Sanders et al., 2002).

The results of the word discrimination task, in which participants were exposed to trisyllabic combinations that had occurred with different frequencies within the speech streams, provide additional evidence that word segmentation had implicitly occurred during the speech stream exposure task. Here, greater

Table 7. Peaks of activation in word discrimination task

Anatomical regions	BA		Word > nonword				Word > partword				Partword > nonword			
			x	*y*	*z*	*t*	*x*	*y*	*z*	*t*	*x*	*y*	*z*	*t*
Inferior frontal gyrus	44	L	−54	8	22	4.01	−44	16	22	5.65				
	45	L	−50	20	20	4.08								
Middle frontal gyrus	9	L	−44	24	30	5.29	−44	26	32	4.12	−44	24	26	4.11
Inferior parietal lobule	40	R									−46	−40	32	5.04

Activity thresholded at *t* > 3.30 (*p* < 0.001), corrected for multiple comparisons at the cluster level (*p* < 0.05). L, Left; R, right.

activity was observed in left prefrontal regions in response to trisyllabic combinations with higher frequency of occurrence within the speech streams and, thus, higher transitional probabilities between syllables (Fig. 7; Table 7). Specifically, listening to words compared with partwords (i.e., trisyllabic combinations that had occurred 45 times versus 15 times within the artificial language streams) and nonwords (i.e., trisyllabic combinations that occurred only once within the random syllables stream) elicited greater activity in the posterior, superior region of left IFG extending into the MFG, regions known to be important for phonological and sequential processing (Gelfand and Bookheimer, 2003). That this effect was more pronounced for the words from the stressed language than the unstressed language again suggests that participants were able to capitalize on the presence of prosodic cues to aid word segmentation during the speech stream exposure task. The observed left-lateralized prefrontal activity is consistent with a host of previous imaging studies implicating these regions in several aspects of phonological processing (for review, see Bookheimer, 2002), including phonemic discrimination, temporal sequencing, articulatory recoding, and the maintenance of acoustic information in working memory (Burton et al., 2000; Newman and Twieg, 2001; Gelfand and Bookheimer, 2003; LoCasto et al., 2004; Xiao et al., 2005).

Interestingly, activity within the left MFG while listening to words from within the stressed and unstressed languages was modulated by participants' rapid auditory processing skills, as indexed by their performance on the Tallal Repetition Test (Tallal and Piercy, 1973) (Fig. 8). Previous research has shown that the ability to process rapid acoustic changes is a highly significant predictor of future language outcomes and verbal intelligence and that this ability is compromised in individuals with language learning impairments such as dyslexia (Benasich and Tallal, 2002; Tallal, 2004). The observed relationship between MFG activity (taken to index successful implicit word segmentation) and individual differences in rapid auditory processing skills further attests to the importance of this region for language learning. Although other neuroimaging studies have linked MFG activity to the processing of rapid acoustic information at the phonemic level (Fiez et al., 1995; Temple et al., 2000; Poldrack et al., 2001; Temple, 2002), the present results indicate that rapid auditory processing skills may also be important for other aspects of language learning.

In conclusion, the current research confirms that minimal exposure to a continuous stream of speech containing statistical and prosodic cues is sufficient for implicit word segmentation to occur and provides a neural signature of on-line word segmentation in the mature brain. At an applied level, this paradigm may be useful for identifying abnormalities in the neural architecture subserving language learning in developmental language disorders and for exploring changes in this circuitry after interventions. At a theoretical level, given that infants and adults can segment sequences of tones just as well as syllables (Saffran et al., 1999), a comparison of neural activity during speech parsing versus parsing of nonlinguistic stimuli such as tones may allow us to disentangle domain-specific from domain-general mechanisms subserving language learning. Most importantly, the present findings in adults can serve as a neurodevelopmental endpoint to investigate changes in the neural basis of language learning occurring with age and linguistic experience to help answer the long-standing question of why children are better language learners than adults.

References

Aslin RN, Saffran JR, Newport EL (1998) Computation of probability statistics by 8-month-old infants. Psychol Sci 9:321–324.

Benasich A, Tallal P (2002) Infant discrimination of rapid auditory cues predicts later language impairment. Behav Brain Res 136:31–49.

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

Bischoff-Grethe A, Proper SM, Mao H, Daniels KA, Berns GS (2000) Conscious and unconscious processing of nonverbal predictability in Wernicke's area. J Neurosci 20:1975–1981.

Blakemore SJ, Rees G, Frith CD (1998) How do we predict the consequences of our actions? A functional imaging study. Neuropsychologia 36:521–529.

Bookheimer S (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151–188.

Burton MW, Small SL, Blumstein SE (2000) The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci 12:679–690.

Callan AM, Callan DE, Masaki S (2005) When meaningless symbols become letters: neural activity change in learning new phonograms. NeuroImage 28:553–662.

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

Casey BJ, Galvan A, Hare TA (2005) Changes in cerebral functional organization during cognitive development. Curr Opin Neurobiol 15:239–244.

Cutler A, Carter DM (1987) The predominance of strong initial syllables in the English vocabulary. Comput Speech Lang 2:133–142.

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S (2005) Neural correlates of switching from auditory to speech perception. NeuroImage 24:21–33.

Doyon J, Penhune V, Ungerleider LG (2003) Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. Neuropsychologia 41:252–262.

Fiez JA, Raichle ME, Miezin FM, Petersen SE, Tallal P, Katz WF (1995) PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands. J Cogn Neurosci 7:357–375.

Fletcher PC, Zafiris O, Frith CD, Honey RAE, Corlett PR, Zilles K, Fink GR (2004) On the benefits of not trying: brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. Cereb Cortex 15:1002–1015.

Friederici AD, Steinhauer K, Pfeifer E (2002) Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. Proc Natl Acad Sci USA 99:529–534.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. NeuroImage 10:385–396.

Gandour J, Wong D, Lowe M, Dzemidzic M, Satthamnuwong N, Tong Y, Li X (2002) A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. J Cogn Neurosci 14:1076–1087.

Gelfand J, Bookheimer SY (2003) Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron 38:831–842.

Golestani N, Zatorre RJ (2004) Learning new sounds of speech: reallocation of neural substrates. NeuroImage 21:494–506.

Hashimoto R, Sakai KL (2004) Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. Neuron 42:311–322.

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

Jaeggi SM, Seewer R, Nirkko AC, Eckstein D, Schroth G, Groner R, Gutbrod K (2003) Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. NeuroImage 19:210–225.

Jäncke L, Wustenberg T, Scheich H, Heinze HJ (2002) Phonetic perception and the temporal cortex. NeuroImage 15:733–746.

Johnson EK, Jusczyk PW (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. J Mem Lang 44:548–567.

Johnson EK, Newport EL (1989) Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. Cognit Psychol 21:60–99.

Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123:155–163.

Jusczyk PW (2002) How infants adapt speech-processing capacities to native-language structure. Curr Dir Psychol Sci 11:15–18.

Kelly AMC, Garavan H (2005) Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15:1089–1102.

Kuhl PK (2004) Early language acquisition: cracking the speech code. Nat Rev Neurosci 5:831–843.

Landau SM, Schumacher EH, Garavan H, Druzgal TJ, D'Esposito (2004) A functional MRI study of the influence of practice on component processes of working memory. NeuroImage 22:211–221.

LoCasto PC, Krebs-Noble D, Gullapalli RP, Burton MW (2004) An fMRI investigation of speech and tone segmentation. J Cogn Neurosci 16:1612–1624.

McLaughlin J, Osterhout L, Kim A (2004) Neural correlates of second-language word learning: minimal instruction produces rapid change. Nat Neurosci 7:703–704.

Newman SD, Twieg D (2001) Differences in auditory processing of words and pseudowords: an fMRI study. Hum Brain Mapp 14:39–47.

Noppeney U, Price CJ (2004) An fMRI study of syntactic adaptation. J Cogn Neurosci 16:702–713.

Opitz B, Friederici AD (2003) Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. NeuroImage 19:1730–1737.

Opitz B, Friederici AD (2004) Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. J Neurosci 24:8436–8440.

Petersen SE, van Mier H, Fiez JA, Raichle ME (1998) The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci USA 95:853–860.

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JDE (2001) Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J Cogn Neurosci 13:687–697.

Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, Knowlton BJ (2005) The neural correlates of motor skill automaticity. J Neurosci 25:5356–5364.

Saffran JR, Wilson DP (2003) From syllables to syntax: multilevel statistical learning by 12-month-old infants. Infancy 4:273–284.

Saffran JR, Aslin RN, Newport EL (1996a) Statistical learning by 8-month-old infants. Science 274:1926–1928.

Saffran JR, Newport EL, Aslin RN (1996b) Word segmentation: the role of distributional cues. J Mem Lang 35:606–621.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70:27–52.

Saint-Cyr JA (2003) Frontal-striatal circuit functions: context, sequence, and consequence. J Int Neuropsychol Soc 9:103–128.

Sanders LD, Newport EL, Neville HJ (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat Neurosci 5:700–703.

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100–107.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shadmehr R, Holcomb HH (1997) Neural correlates of motor memory consolidation. Science 277:821–825.

Shestakova A, Huotilainen M, Ceponiene R, Cheour M (2003) Event-related potentials associated with second language learning in children. Clin Neurophysiol 114:1507–1512.

Tallal P (2004) Improving language and literacy is a matter of time. Nat Rev Neurosci 5:721–728.

Tallal P, Piercy M (1973) Defects of non-verbal auditory perception in children with developmental dysphasia. Nature 241:468–469.

Temple E (2002) Brain mechanisms in normal and dyslexic readers. Curr Opin Neurobiol 12:178–183.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JDE (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proc Natl Acad Sci USA 97:13907–13912.

Thiel CM, Shanks DR, Henson RNA, Dolan RJ (2003) Neuronal correlates of familiarity-driven decisions in artificial grammar learning. NeuroReport 14:131–136.

Thiessen ED, Saffran JR (2003) When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Dev Psychol 39:706–716.

Van der Graaf FHCE, de Jong BM, Maguire RP, Meiners LC, Leenders KL (2004) Cerebral activation related to skills practice in a double serial reaction time task: striatal involvement in random-order sequence learning. Cognit Brain Res 20:120–131.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001) Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci 13:994–1005.

Weber-Fox C, Neville HJ (2001) Sensitive periods differentiate processing of open- and closed-class words: an ERP study of bilinguals. J Speech Lang Hear Res 44:1338–1353.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702.

Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC (1998a) Automated image registration. I. General methods and intrasubject, intramodality validation. J Comput Assist Tomogr 22:139–152.

Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC (1998b) Automated image registration. II. Intersubject validation of linear and nonlinear models. J Comput Assist Tomogr 22:153–165.

Woods RP, Dapretto M, Sicotte NL, Toga AW, Mazziotta JC (1999) Creation and use of a Talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. Hum Brain Mapp 8:73–79.

Xiao Z, Zhang JX, Wang X, Wu R, Hu X, Weng X, Tan LH (2005) Differential activity in left inferior frontal gyrus for pseudowords and real words: an event-related fMRI study on auditory lexical decision. Hum Brain Mapp 25:212–221.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.