Published in final edited form as:

Dev Sci. 2010 March ; 13(2): 385–406. doi:10.1111/j.1467-7687.2009.00895.x.

The Neural Basis of Speech Parsing in Children and Adults

Kristin McNealy^{1,5,6}, John C. Mazziotta^{1,3,4,5}, and Mirella Dapretto^{1,2,3,5,6}

¹Ahmanson-Lovelace Brain Mapping Center, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, 90095, USA

²Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, California, 90095, USA

³The Brain Research Institute, University of California, Los Angeles, California, 90095, USA

⁴Departments of Neurology, Pharmacology, and Radiological Sciences in the David Geffen School of Medicine, University of California, Los Angeles, California, 90095, USA

⁵Neuroscience Interdepartmental Program, University of California, Los Angeles, California, 90095, USA

⁶Center for Culture, Brain, and Development, University of California, Los Angeles, California, 90095, USA

Abstract

Word segmentation, detecting word boundaries in continuous speech, is a fundamental aspect of language learning that can occur solely by the computation of statistical and speech cues. Fifty-four children underwent functional magnetic resonance imaging (fMRI) while listening to three streams of concatenated syllables, which contained either high statistical regularities, high statistical regularities and speech cues, or no easily-detectable cues. Significant signal increases over time in temporal cortices suggest that children utilized the cues to implicitly segment the speech streams. This was confirmed by the findings of a second fMRI run where children displayed reliably greater activity in left inferior frontal gyrus when listening to ‘words’ that occurred more frequently in the streams of speech they just heard. Finally, comparisons between activity observed in these children vs. previously-studied adults indicate significant developmental changes in the neural substrate of speech parsing.

Keywords

fMRI; language; development; speech perception; word segmentation; statistical learning

Rapid and continuous developments in non-invasive imaging techniques witnessed in the last decade have allowed for significant strides to be made in delineating the functional representation of language in the developing brain (e.g., Booth, Cho, Burman, & Bitan, 2007; Brown et al., 2005; Holland et al., 2001; Plante, Holland, & Schmithorst, 2006; Schlaggar, Brown, Lugar, Visscher, Miezin, & Petersen, 2002; Turkeltaub, Gareau, Flowers, Zeffiro, & Eden, 2003; and see Friederici, 2006 for a review); however, we still know very little about the neural circuitry involved in the process of language learning in children. Converging evidence from cleverly designed behavioral studies shows that key components of language acquisition involve the extraction of patterns from a number of probabilistic

cues available in the input (see Kuhl, 2004 for a review). Here, we investigated the mechanisms by which the developing brain encodes the statistical regularities and prosodic cues present in continuous speech to identify word boundaries (i.e., word segmentation). Sensitivity to statistical regularities in the input has been demonstrated to underlie sequence learning across visual, auditory, and motor domains in both human and nonhuman primates (e.g., Fiser & Aslin, 2002; Hauser, Newport, & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002; Robertson, 2007; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). Importantly, with regard to the aims of the present research, statistical learning has been shown to play a role in different aspects of language learning in addition to word segmentation (e.g., Saffran et al., 1996), including phonetic learning (Kuhl et al., 2006; Maye, Werker, & Gerken, 2002), artificial grammar learning (Opitz & Friederici, 2003, 2004), acquiring knowledge of verb-argument structure (Wonnacott, Newport, & Tanenhaus, 2007), as well as the learning of word-referent mappings (Smith & Yu, 2007).

Thus far, several approaches have been used to examine language learning in the adult brain with functional magnetic resonance imaging (fMRI). Some studies have focused on the neural reorganization that results from a period of intensive training on a novel linguistic task (e.g., Callan et al., 2003; Callan, Callan, & Masaki, 2005; Friederici, Steinhauer, & Pfeifer, 2002; Golestani & Zatorre, 2004; Newman-Norlund, Frey, Petitto, & Grafton, 2006). Other investigations have explored the neural basis of the language learning process as it actually occurs by examining changes in brain activity while participants visually learn an artificial grammar in the scanner (e.g., Hashimoto & Sakai, 2004; Opitz & Friederici, 2003, 2004; Thiel, Shanks, Henson, & Dolan, 2003). To our knowledge, only one fMRI study has evaluated learning-associated changes in neural activity during initial language learning as it occurs in the auditory modality, as adults listened to artificial languages and began to identify word boundaries in the continuous speech streams (McNealy, Mazziotta, & Dapretto, 2006).

The identification of word boundaries in continuous speech is a crucial aspect of language learning that must take place before any further linguistic analysis can be performed (e.g., Jusczyk, 2002; Saffran & Wilson, 2003). Infants as young as 7.5 months of age have been shown to compute information about the distributional frequency with which certain syllables occur in relation to others and calculate the odds (transitional probabilities) that one syllable will follow another in a given language to successfully parse a continuous stream of speech (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). 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 & Jusczyk, 2001; Thiessen & Saffran, 2003). Importantly, the ability to discriminate words from fluent speech in infancy has been linked to higher vocabulary scores at two years of age and better overall language skills in preschool (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006). Furthermore, evidence of a direct connection between speech parsing and learning the meaning of words comes from a recent study showing that 17-month old infants' learning of object labels is facilitated by prior experience with segmenting a continuous speech stream that contained the novel object names as words (Graf Estes, Evans, Alibali, & Saffran, 2007).

Given the fundamental role that speech parsing plays in language learning, we previously adapted a well-established word segmentation paradigm from the infant behavioral literature to explore how the adult brain processes statistical and prosodic cues in order to segment continuous speech streams (McNealy et al., 2006). While undergoing fMRI, participants first listened to three distinct streams of concatenated syllables, containing either strong statistical cues to word boundaries, strong statistical and speech cues, or no easily detectable cues (Speech Stream Exposure Task). Listening to each of the three speech streams engaged

bilateral fronto-temporal networks; however, left-lateralized signal increases *over time* were observed in temporal and inferior parietal regions only when adults listened to the artificial language streams containing high statistical regularities, a finding which suggests that they were able to implicitly segment these streams by computing the statistical relationships between syllables online. Immediately following exposure to the speech streams, participants underwent a second fMRI scan (Word Discrimination Task) where they listened to trisyllabic combinations that occurred with different frequencies in the streams of speech they just heard (“words,” 45 times; “partwords,” 15 times; and “nonwords,” once). Reliably greater activity in left inferior and middle frontal gyri was observed in these adult participants when comparing words with partwords and, to a lesser extent, when comparing partwords with nonwords, supporting the notion that implicit speech parsing had taken place. It should be noted that, although adults were able to discriminate at the neural level between words and partwords, this study did not directly examine whether they extracted word-like units from the speech streams. However, the findings of Graf Estes and colleagues (2007), have demonstrated that the ultimate outcome of tracking statistical regularities in the input is successful word segmentation. Hence, while we refer throughout this manuscript to implicit word segmentation, we acknowledge that the listeners’ degree of learning, particularly after such a short period of exposure, may not have proceeded as far as to extract word-like units.

The present research utilized the same fMRI paradigms to examine neural activity related to speech parsing in ten-year-old children, before the time when significant decrements in the ability to fully master a new language are typically observed (e.g., Johnson & Newport, 1989; Weber-Fox & Neville, 2001). In order to characterize developmental changes in the neural architecture underlying language learning, we directly compared neural activity observed in the present sample of children to that previously observed in adults (McNealy et al., 2006). We expected that children would engage bilateral fronto-temporal language networks while listening to the speech streams during the Speech Stream Exposure Task, as did adults; however, based on studies reporting developmental changes in which task-related activation becomes less diffuse and more focal over time (e.g., Casey, Galvan, & Hare, 2005; Durston & Casey, 2005), we predicted that children may recruit a larger overall network than adults while listening to each speech stream. We also reasoned that if children are more sensitive to small transitional probabilities than adults, signal increases over time in temporal and inferior parietal regions could be observed not only as a function of exposure to the speech streams containing high statistical regularities and prosodic cues (as was observed in adults), but also to the stream where statistical cues to word boundaries are minimal (i.e., transitional probabilities between syllables average 0.1 as opposed to 0.76 in the streams with high statistical regularities).

For the Word Discrimination Task, we hypothesized that the tracking of statistical regularities during exposure to the speech streams would result in stronger activity in response to words compared to partwords and nonwords, as children recognize those trisyllabic combinations as having occurred together more often, and potentially as having “word-like” status. Further, we reasoned that if children’s advantage in learning new languages with native-like proficiency might be related to an enhanced ability to segment words within continuous speech that declines with age, children might display greater differential activity between words, partwords and nonwords than adults. However, given that English phonemes and syllables were used to form the speech streams and that adults’ experience in processing them far exceeds that of children, we also considered that adults might be better able to segment the speech streams than children, in which case the opposite pattern of results might be observed.

The use of this word segmentation paradigm in an fMRI study builds upon previous behavioral research, which critically revealed the importance of statistical computations for sequence learning and language acquisition. First, this study allows us to examine which brain regions are involved in performing these statistical computations during language learning. Second, it permits us to identify developmental changes in the neural mechanism of speech parsing, even though behavioral differences between children and adults are not apparent, given that both groups perform at chance unless prolonged exposure and/or explicit training are provided (Saffran, Newport, & Aslin, 1996; Sanders, Newport, & Neville, 2002). Also, the behavioral assessment of the ability of older children and adults to discriminate between words and partwords differs considerably from the infants' head turn novelty preference procedure. Accuracy and reaction time data are important for determining whether an individual can discriminate between words and partwords, but these are measures of the outcome of learning. The fMRI data in the current study can provide insight into the neural mechanism of the learning process itself during exposure to the streams of speech (speech stream exposure task), and can further reveal discrimination between words, partwords, and nonwords at the neural level (word discrimination task). The fMRI data thus serves as a new dependent measure that can be used to inform us about how and where the brain processes statistical and prosodic cues during initial, implicit language learning.

Materials and Methods

Participants

Fifty-four typically developing children (27 female; mean age, 10.1 years, range, 9.5–10.6 years) were recruited from the greater Los Angeles area via summer camps, posted flyers, and mass mailings. Written informed child assent and parental consent was obtained from all children and their parents, respectively, according to the guidelines set forth by the UCLA Institutional Review Board. Children's parents filled out several questionnaires that assessed their child's handedness, health history, and language background. All children were right-handed and had no history of significant medical, psychiatric, or neurological disorders on the basis of parental reports on a detailed medical questionnaire, the Child Behavior Checklist (Achenbach & Edelbrock, 1983), and a brief neurological exam (Quick Neurological Screening Test II; Mutti, Sterling, Martin, & Spalding, 1998). All children were native English speakers, except for three children who learned English before the age of 5 and for whom English was the language in which they are most proficient.

All 54 children completed an fMRI speech stream exposure scan and 12 of them (6 female; mean age, 10.1 years) also underwent an event-related fMRI word discrimination scan immediately after the speech stream exposure scan. All children completed a behavioral word discrimination task outside of the scanner following the scanning session.

Data from this sample were compared with those obtained from a sample of 27 adults as previously reported (McNealy et al., 2006). By report, these adults (13 female; mean age, 26.63 years; range, 20 – 44 years) were right-handed, native English speakers with no history of neurological or psychiatric disorders.

Stimuli and Tasks

Speech stream exposure task—Children listened to three counterbalanced streams of nonsense speech supposedly spoken by aliens from three different planets. Children 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 1A–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 4 trisyllabic words, following the exact same procedure used in previous infant and adult behavioral studies (see Saffran et al., 1996; Saffran, Newport, & Aslin, 1996; Aslin et al., 1998; Johnson & Jusczyk, 2001). Each syllable was recorded separately using SoundEdit, ensuring that the average syllable duration (0.267 sec), amplitude (18.2 dB), and pitch (221 Hz) were (1) not significantly different across the experimental conditions and (2) matched to those previously used in the behavioral literature. For each artificial language, the 4 words were randomly repeated 3 times to form a block of 12 words, subject to the constraint that no word repeated twice in a row. Five such different blocks were created, and then this 5-block sequence was itself concatenated 3 times to form a continuous speech stream lasting 2 minutes and 24 seconds, 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 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 sec), 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 ninety percent 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 facilitate a comparison of activity associated with processing input with high statistical and prosodic cues to word boundaries, as in the two artificial language conditions (U+S), and activity related to listening to a series of concatenated syllables that cannot be readily parsed into words. In this condition, the twelve 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, the statistical likelihood of any one syllable following another was very low (with an average transitional probability between syllables in the stream of 0.1; range 0.02 – 0.22). While a listener may attempt to calculate the transitional probabilities between syllables in this stream, no reliable statistical cues nor any prosodic cues were afforded to the listener to aid speech parsing. Thus, as depicted in Figure 1C, each child listened to three 144-second speech streams (R, U, and S) interspersed between 30 seconds of resting baseline. Short samples of these speech streams are available in Supplementary Materials. The order of presentation of the three experimental conditions was counterbalanced across children according to a Latin Square design.

It is important to note that: 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 children, each set of twelve syllables was used with the same frequency in each condition, thus ensuring that any difference between conditions would not be due 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 prior 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 children); and 4) while 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 our previous word segmentation fMRI study in adults demonstrating that sufficient power was nevertheless achieved (McNealy et al., 2006).

Word discrimination task—Twelve children also completed a second fMRI task to provide an index of implicit 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, children 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 prior behavioral studies, children 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, Newport, et al., 1996; Sanders, Newport, & Neville, 2002). Accordingly, children 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 2A–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 previously heard by a participant. Within each block, children listened to the 4 “words” (W) used to create the artificial language streams (e.g., “golatu”, “daropi”) as well as to 4 “partwords” (PW), that is, trisyllabic combinations 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 “...golat**u**daropi...”). 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), due to the words and partwords having occurred within the speech stream 45 and 15 times, respectively. Each word and partword was repeated five times and interspersed with null events. The same pseudo-randomized 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 four 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 Syllables condition effectively served as “nonwords” (NW), since 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).

It is worth mentioning that, because the overall frequency of words and partwords differ in the speech stream exposure task, this design does not allow for the distinction to be made as to whether the participants are calculating transitional probabilities or frequency of co-occurrence between adjacent syllables. The results of a previous behavioral study designed precisely to discern which type of statistical computations learners perform indicate that learners do track transitional probabilities (Aslin et al., 1998). Hence, when we refer to calculation of statistical regularities, we mean this to include the tracking of both frequency of co-occurrence between adjacent syllables and transitional probabilities.

Behavioral task—To investigate whether children 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. Children 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 had previously heard. (Responses from one child were not recorded due to a computer malfunction.)

fMRI Data Acquisition

Functional images were collected using a Siemens Allegra 3 Tesla head-only MRI scanner. A 2D spin-echo scout (TR = 4000 ms, TE = 40 ms, matrix size 256 by 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 child, a high-resolution structural T2-weighted echo-planar imaging volume (spin-echo, TR = 5000 ms, TE = 33 ms, matrix size = 128 by 128, FOV = 20cm, 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 child's data into a standard coordinate system. For the Speech Stream Exposure Task, one functional scan lasting 8 minutes and 48 seconds was acquired covering the whole cerebral volume (174 images, EPI gradient-echo, TR = 3000 ms, TE = 25 ms, flip angle = 90°, matrix size = 64 by 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 9 minutes and 18 seconds was acquired that also covered the whole brain (277 images, EPI gradient-echo, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix size = 64 by 64, FOV = 20 cm, 36 slices, 3.125-mm in-plane resolution, 3-mm thick, 1-mm gap).

Children listened to the auditory stimuli through a set of magnet-compatible stereo headphones (Resonance Technology, Inc.). Stimuli were presented using MacStim 3.2 software (Darby, WhiteAnt Occasional Publishing and CogState Ltd, 2000).

fMRI Data Analysis

Using Automated Image Registration (AIR; Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998a; Woods, Grafton, Watson, Sicotte, & Mazziotta, 1998b) functional images for each child were 1) realigned to each other to correct for head motion during scanning and co-registered 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 Talairach-compatible MR atlas (Woods, Dapretto, Sicotte, Toga, & Mazziotta, 1999) using polynomial non-linear warping; and 3) smoothed with a 6mm FWHM 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 child, 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, Holmes, Price, Buchel, & Worsley, 1999). For all comparisons vs. resting baseline (both within- and between-group), 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). As noted in the tables below, for contrasts examining signal increases over time, or the relationship between activation and behavioral indices of word segmentation, activity was deemed reliable if it survived a slightly less stringent magnitude threshold of $t > 2.40$ ($p < 0.01$, uncorrected) or small volume correction (5mm minimum sphere radius), provided that this

activity fell within a priori regions of interest, that is, regions where reliable activity was observed in our earlier study in adults (McNealy et al., 2006). Small volume correction at the cluster level was also used in the basal ganglia.

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. Further, because statistical regularities are computed online during the course of listening to the speech streams, we also examined whether children, like adults, would display signal increases over time in language-relevant cortices as a function of exposure to the speech streams. A region-of-interest analysis was then conducted in bilateral temporal cortices where such signal increases were observed in order to examine whether there were any hemispheric laterality differences. This functionally-defined ROI included all voxels showing reliable signal increases at the group level in either the left or right hemisphere (LH and RH), as well as their respective counterparts in the opposite hemisphere. For each child, 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).

To assess whether activity observed during exposure to the artificial languages might reflect successful implicit word segmentation resulting from the computation of statistical regularities and prosodic cues, regression analyses were also conducted to assess the relationship between accuracy scores and reaction times on the behavioral post-scanner Word Discrimination test and neural activity associated with listening to the two artificial languages, as compared to 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 in order to test the hypothesis that children, like adults, would display differential activity in language networks for Words, Partwords, and Nonwords.

For both the Speech Stream Exposure Task and the Word Discrimination Task, two-sample *t* tests were implemented to compare neural activity observed in the current sample of children to those observed in our previous study of word segmentation in adults (McNealy et al., 2006). The SPM toolbox MarsBaR (<http://marsbar.sourceforge.net>) was used to extract parameter estimates for each participant from regions that were significantly active in the between group comparison of the two artificial languages and the Random Syllables stream. Importantly, no between-group differences were found in the mean amount of head motion for either scan.

Results

Behavioral Results

Accuracy and response times on the behavioral word recognition test conducted after the fMRI scans are reported in Table 1. In line with the behavioral findings of prior word segmentation studies (e.g., McNealy et al., 2006; Saffran, Newport, et al., 1996; Sanders et al., 2002), children 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 in any condition. Similarly, there were no significant differences between conditions in response times to either words or partwords, although responses to words were overall significantly faster than responses to partwords [$F(1,51) = 15.4, p < 0.001$]. When comparing the behavioral responses of children and adults, there were no significant group differences in accuracy [$F(1,77) = 0.017, ns$], but adults were overall significantly faster when responding to both words and partwords [$F(1,77) = 35.5, p < 0.001$].

Speech Stream Exposure Task

As expected and similar to what was previously observed in adults, merely listening to all three streams of continuous speech (vs. resting baseline) activated a large-scale bilateral neural network including canonical language areas in the left hemisphere (LH) as well as extensive bilateral activation in temporal, parietal and frontal cortices (Figure 3 and Table 2). In addition, when children listened to the Stressed (S) and Unstressed (U) Language streams, bilateral activity in the basal ganglia – subcortical structures involved in implicit learning – was also detected (Talairach coordinates, $-22, 2, 8, t = 3.33$, and $22, -4, 8, t = 3.13$). When children listened to the Random Syllables (R) stream, additional activity was observed in the right hemisphere counterpart of Broca's area (inferior frontal gyrus, IFG; $48, 6, 16, t = 3.14$), in line with prior evidence that language-related activity in this region increases as a function of task difficulty.

Direct comparisons between conditions were then implemented to examine whether children would exhibit, as did adults, greater activity when attempting to parse a stream of speech with low statistical regularities compared to streams of speech with high statistical regularities. These analyses indeed revealed greater activity in several areas when listening to the stream of random syllables, as compared to the artificial languages [(R – U) and (R – S)], in several premotor regions often activated during language tasks although outside of what has been traditionally considered language networks (Table 3). In contrast, when comparing activity while listening to the artificial languages versus the random syllables, significantly greater activity was observed in two canonical LH language areas (left inferior parietal lobule, IPL, and superior temporal gyrus, STG) for the Unstressed Language (U – R; Table 3), and in the right STG, a region implicated in the processing of speech prosody, for the Stressed Language (S – R; Table 3). To further interrogate differential activity related to the presence of prosodic cues within the artificial language stream, neural activity associated with listening to the Stressed Language was directly compared to that associated with listening to the Unstressed Language. For this comparison (S – U), greater activity was also observed in the right STG ($54, -22, 6, t = 3.78$), as well as in the anterior cingulate gyrus (CG; $22, 16, 28, t = 3.90$) and supplementary motor area (SMA, $12, 8, 60, t = 3.39$), suggesting that the presence of prosodic cues modulated activity within both language and attentional networks. As could be expected given that the Unstressed and Stressed languages shared the same statistical properties, the opposite comparison (U – S) revealed no significantly greater activity for the Unstressed Language.

Because statistical regularities are computed online during the course of the activation conditions, regions where activity might be increasing as a function of exposure to the speech streams were examined next in order to detect 'learning-related' signal increases over time. A statistical contrast modeling increases in activity over time for the two artificial languages ($U\uparrow + S\uparrow$) revealed significant bilateral signal increases in STG and transverse temporal gyrus (TTG), extending into the IPL in the LH, as well as in bilateral putamen and left SMA, a pattern that is similar to what was observed in adults (Figure 4 and Table 4). Unlike the prior findings in adults, however, a region of interest analysis indicated that this increasing activity during exposure to the Unstressed and Stressed Language conditions was

not lateralized to the LH, as activity did not significantly differ between the LH and RH [$t(53) = 0.63$, *ns*]. Importantly, whereas adults displayed no reliable signal increases over time during exposure to the Random Syllables stream, children did display some significant signal increases in bilateral STG for this condition (Figure 4 and Table 4), evidence that lends support to the hypothesis that children might be better than adults at tracking statistical regularities in sequences with low frequency of co-occurrence between syllables (Figure 4 and Table 4).

To characterize the differential strength of the signal increases over time for each condition, we directly compared signal increases occurring during exposure to the two artificial languages to the signal increases occurring during exposure to the Random Syllables condition ($(U\uparrow + S\uparrow) - R\uparrow$). This statistical contrast revealed significantly greater signal increases for the artificial languages in left STG and IPL (Table 4). It may be worth noting that this pattern held when comparing signal increases during each artificial language condition to the Random Syllables condition ($U\uparrow - R\uparrow$ and $S\uparrow - R\uparrow$; Table 5). Unexpectedly, in light of the prior findings in adults that showed stronger signal increases for the Stressed Language (vs. the Unstressed one), direct comparisons between the two artificial languages in children revealed stronger signal increases in bilateral STG (Table 5) for the Unstressed Language versus Stressed Language ($U\uparrow - S\uparrow$). There were no regions where activity was increasing to a greater extent during exposure to the Random Syllables condition than to the artificial language conditions. In addition, a statistical contrast modeling decreases in activity over time revealed that there were no significant signal decreases for any of the three speech streams ($R\downarrow$, $U\downarrow$, and $S\downarrow$).

Comparison between Children and Adults – Speech Stream Exposure Task

In order to assess developmental differences in the neural substrate of speech parsing, we directly compared patterns of activity between children and adults. When we examined group differences in activity for each of the three individual speech streams (vs. resting baseline), we observed a large degree of similarity across these three comparisons. Hence, we describe the overall effects below and present more detailed information about each contrast in Tables 6 and 7. As predicted from reports of more diffuse task-related activation in other developmental fMRI studies (e.g., Casey et al., 2005; Durston & Casey, 2005), children showed greater activity than adults in left IPL and superior parietal lobule (SPL), right PCL, and bilateral precuneus (PCU; Figure 5; Table 6). This finding suggests that children recruit a more dorsal and posterior attentional network to process the statistical and prosodic cues available in the speech streams. Children also displayed greater activity than adults in bilateral SMA and the left basal ganglia while listening to the Stressed Language. In turn, significantly greater activity was observed for adults compared to children in medial prefrontal cortex as well as in canonical language regions (Figure 5; Table 7), including bilateral IFG, STG, and middle temporal gyrus (MTG).

We next examined whether there were any reliable group differences between conditions, when adults and children listened to the artificial language streams, which contained high statistical regularities as cues to word boundaries, compared to the Random Syllables stream, which contained only minimal statistical regularities. The results revealed reliable interaction effects in several regions, including left inferior frontal and superior temporal cortices and bilateral middle temporal and inferior parietal cortices, paracentral lobules (PCL) and putamen (Table 8). To qualify these results, we extracted parameter estimates of activity from the maxima in each of these regions for children and adults (using MarsBaR). An examination of these parameter estimates provided confirmation that these interaction effects were attributable to significantly greater activation for adults than children for the Random Syllables stream as compared to the artificial language streams [$R - (U + S)$]. This finding is consistent with the between-condition results within each group, whereby adults

displayed greater activity in many regions for the Random Syllables condition compared to the artificial language conditions, whereas which children displayed fewer differences between conditions (see description of results for children and adults above and in McNealy et al., 2006, respectively).

Next, differences between children and adults were examined in regions where activity was found to be increasing over the course of listening to the speech streams (i.e., as the statistical regularities were being computed) in order to explore the extent to which recruitment of these regions during statistical learning might change between pre-adolescence and adulthood. The results of these exploratory analyses showed that children displayed significantly greater signal increases than adults during exposure to the artificial language streams ($U\uparrow + S\uparrow$) in left IPL and SMA, whereas adults displayed significantly greater signal increases than children in bilateral STG and MTG (Figure 6; Table 9). Because children displayed significant signal increases during exposure to the Random Syllables stream ($R\uparrow$) whereas adults did not, the between group comparison for this contrast confirmed, as expected, significantly greater signal increases in bilateral STG and left IPL for the children than adults (Figure 6; Table 9).

Correlations with Behavioral Word Discrimination Accuracy and Reaction Time

Regression analyses were performed to explore whether, like adults, children's neural activity while tracking statistical regularities is linked to behavioral accuracy at discriminating between words and partwords. In children, greater accuracy at discriminating words on the post-scan behavioral test was positively correlated with activity in left STG ($-64, -42, 10, t = 3.01$) when listening to the Stressed and Unstressed Languages, as compared to Random Syllables, during the Speech Stream Exposure Task, a finding that is remarkably similar to what was previously observed in adults. When the relationship between children's reaction times for word discrimination and activity while listening to the artificial languages was examined, faster response times were associated with greater activity in bilateral STG ($-60, -20, 2, t = 2.96; 50, 0, -4, t = 3.13$). Importantly, these effects were observed in regions that showed increasing activity over time during exposure to the artificial languages, suggesting that activity in these regions is indeed related to learning.

Word Discrimination Task

To further verify that statistical learning had taken place during the Speech Stream Exposure task, neural activity associated with listening to the Words (W), Partwords (PW) and Nonwords (NW) presented during the three blocks of the Word Discrimination Task was examined next. Activity in response to hearing both Words and Partwords (which had occurred 45 and 15 times, respectively, during the artificial language conditions of the Speech Stream Exposure Task) was summed across the Unstressed and Stressed Language blocks (as was done in our prior study in adults, McNealy et al., 2006) because collapsing across the language blocks yielded 40 events for Words and 40 for Partwords to be contrasted with the 40 events for the Nonwords (i.e., trisyllabic combinations presented during the Random Syllables block which had occurred only once during the Random Syllables condition of the Speech Stream Exposure Task). To test the hypothesis that, like adults, children would be able to discriminate between Words and Nonwords after listening to continuous speech streams containing statistical regularities, we implemented the statistical contrast comparing Words to Nonwords ($W - NW$). As expected, this analysis revealed significantly greater activity for Words in left IFG ($-40, 28, 20, t = 5.2$), similar to what we previously observed in adults, as well as in the right anterior CG ($14, 34, 14, t = 6.2$; Figure 7). This same pattern of results was also observed when the comparison of Words and Nonwords was conducted for the Stressed and Unstressed Language blocks

separately, albeit with smaller cluster size (Stressed W – NW $-40, 28, 20, t = 4.28$, and $22, 30, 20, t = 3.75$, small volume correction applied; Unstressed W – NW $-36, 26, 18, t = 3.42$, and $14, 42, 8, t = 3.79$, small volume correction applied). Reliably greater activity was also observed in left IFG when comparing Partwords (summed across the Stressed and Unstressed blocks) to Nonwords (PW – NW; $-40, 30, 16, t = 4.59$, small volume correction applied). Unlike adults, however, children showed no significant differences in the comparison between Words and Partwords (W–PW). No significant differences in neural activity were observed for any of the reverse statistical comparisons (NW – W, NW – PW, and PW – W).

Lastly, unlike what was observed in adults, no differential activity was found as children listened to Words and Partwords that had occurred in the Stressed Language stream when both prosodic and statistical cues were available to guide word segmentation, as opposed to Words and Partwords that had occurred in the Unstressed Language stream containing only statistical cues ($S(W+PW) - U(W+PW)$) or for the reverse comparison ($U(W+PW) - S(W+PW)$).

Comparison between Children and Adults – Word Discrimination Task

The final set of analyses compared neural activity during the Word Discrimination task between children and adults (Table 10). When contrasting activity associated with listening to Words as compared to Nonwords (W–NW), the right anterior CG was found to be significantly more active in children than adults ($14, 44, 0, t = 3.93$, small volume correction applied), whereas the left IFG and MFG were significantly more active in adults than children (Figure 8). No regions showed reliably greater activity in children than adults when comparing activity associated with listening to Words versus Partwords (W–PW) or Partwords versus Nonwords (PW–NW). In contrast, adults displayed greater activity than children in the left IFG for the comparison between Words and Partwords (W–PW), as well as in the left IPL for the comparison between Partwords and Nonwords (PW–NW). Taken together, these findings highlight again the greater recruitment of language-relevant networks in adults and lend support to the notion that adults might be better able to implicitly segment the speech streams than children, perhaps reflecting the fact that English phonemes and syllables were used to form the speech streams and that adults' experience in processing these stimuli far exceeds that of children.

Discussion

One of the first challenges encountered while learning a new language is to identify word boundaries, as there are actually no breaks or pauses between words that reliably indicate where one word ends and the next begins (e.g., Cole & Jakimik, 1980). Infant behavioral studies of word segmentation conducted over the past decade have forced developmental psychologists to re-conceptualize the process of language acquisition, showing that infants approach this feat computationally (e.g., Aslin et al., 1998; Johnson & Jusczyk, 2001; Saffran et al., 1996; Thiessen & Saffran, 2003); however, how the brain utilizes the statistical and prosodic speech cues available in the input to accomplish various aspects of language learning such as speech parsing remains largely unknown. By adapting the paradigm employed in the infant behavioral research for use with fMRI, McNealy and colleagues (2006) previously identified a neural signature of implicit word segmentation in adults. Given that both infants and adults are capable of speech parsing, we expected that the children in the present study would also utilize the available statistical and prosodic speech cues to implicitly segment the continuous speech streams, and we investigated the extent to which similar neural circuitry underlies speech parsing in children and adults. However, in light of the fact that the ease with which individuals can acquire another language decreases with age (e.g., Johnson & Newport, 1989; Weber-Fox & Neville, 2001), we wanted to

explore whether there might be significant developmental differences in the neural mechanism by which this critical initial stage of language learning takes place.

Speech Stream Exposure Task

When children listened to the streams of concatenated syllables, they activated broad bilateral networks in frontal, temporal and parietal cortices that overlapped considerably with the regions found to be active in the previous fMRI study on word segmentation in adults (Figure 3 and Table 2; McNealy et al., 2006), as might be expected given that prior studies involving the presentation of speech or speech-like stimuli have consistently shown recruitment of similar language-processing circuitry in both adults (see Hickok & Poeppel, 2007 for a review) and children (see Friederici, 2006 for a review). However, significant developmental differences were observed when we directly compared neural activity between groups. When faced with attempting to parse these novel speech streams based solely on the presence of statistical and prosodic cues, adults relied on canonical fronto-temporal language networks to a greater extent than children (Figure 5 and Table 7). Children, in contrast, activated a broader network than adults; in particular, they showed greater activity in dorsal parietal multimodal association areas that are often recruited during the allocation of attention and working memory across a wide variety of tasks (Cavanna & Trimble, 2006; Culham & Kanwisher, 2001; Olesen, Westerberg, & Klingberg, 2004; Osaka, Komon, Morishita, & Osaka, 2007; Figure 5 and Table 6). Children's recruitment of regions that fall outside of canonical language networks (e.g., PCU, SPL) has also been observed in other developmental neuroimaging studies of language processing (Brown et al., 2005; Chou et al., 2006; Schmithorst, Holland, & Plante, 2006). Overall, these findings match well with a developmental progression from a pattern of more diffuse activity to one of more focal and enhanced activity within task-relevant regions that has been noted across many developmental studies, particularly in cases where age-related improvements are observed in the cognitive function being investigated (Casey, Galvan, & Hare, 2005; Casey, Giedd, & Thomas, 2000; Durston & Casey, 2005; Durston et al., 2006). Hence, children's greater engagement of dorsal posterior regions during exposure to the speech streams may reflect children's greater difficulty focusing their attention on the speech streams compared to adults, or the relative immaturity of the children's working memory capacity for phonological information (Gathercole, 1999; Luna, Garver, Urban, Lazar, & Sweeney, 2004).

We next examined differences between the two artificial language streams and the Random Syllables stream in order to characterize how neural activity differs when listening to streams of speech containing high statistical regularities that can guide speech parsing as opposed to a stream of syllables with low statistical regularities that could not easily be parsed into words. In our prior study in adults, the presence of high statistical regularities and prosodic cues for the artificial languages facilitated more efficient processing and resulted in more focal activity in many regions for the Unstressed and Stressed Language streams compared to the Random Syllables stream. This trend was not observed to the same degree in children, who only exhibited greater activity while listening to the stream of Random Syllables compared to the artificial languages in bilateral supplementary motor and dorsal premotor cortices, regions which have been shown to be more engaged during speech perception tasks when processing difficulty is higher (Table 3; Wilson & Iacoboni, 2006; Wilson, Molnar-Szakacs, & Iacoboni, 2007). Whereas there were no regions for which adults exhibited greater activity while they listened to the artificial language streams compared to the random syllables stream, children exhibited greater activity for the Unstressed Languages in left superior temporal and inferior parietal cortices and left basal ganglia, as well as greater activity for the Stressed Languages in right superior temporal cortex (Table 3). It should be noted, however, that while the within-group comparisons

between the speech streams yielded different results for children and adults, the significant interaction effects between groups and conditions (observed in bilateral temporal and parietal cortices and left prefrontal regions) were all actually driven by greater activity in adults for the Random stream as compared to the artificial language streams (Table 8).

After characterizing the pattern of activity averaged across each exposure block when participants listened to the speech streams, we then examined where activity might be changing over time as a function of exposure to the speech streams in order to identify activity that might specifically reflect the computation of statistical regularities and the likely implicit detection of word boundaries. Previous studies that entailed the learning of a linguistic task have reported a combination of learning-related increases and decreases in activity (Thiel et al., 2003; Golestani & Zatorre, 2004; Newman-Norlund et al., 2006; Noppeney & Price, 2004; Rauschecker, Pringle, & Watkins, 2007), with increases in activation found in primary and secondary sensory and motor areas, as well as in regions involved in the storage of task-related cortical representations (see Kelly and Garavan, 2005, for a review). Here, BOLD signal increases over time (i.e., within each exposure block) were observed in bilateral temporal cortices and left supramarginal gyrus when children listened to the Stressed and Unstressed Language streams, just as we previously found in adults (Figure 4 and Table 4). However, while adults' signal increases were significantly left-lateralized, the significant signal increases observed in children did not differ between hemispheres. As we previously proposed (McNealy et al., 2006), these signal increases over time along the superior temporal gyri may reflect the ongoing computation of frequencies of syllable co-occurrence and transitional probabilities between neighboring syllables, in line with other studies have shown the STG to be involved in processing the predictability of sequences over time (Blakemore, Rees, & Frith, 1998; Bischoff-Grethe, Proper, Mao, Daniels, & Berns, 2000; Ullen, Bengtsson, Ehrsson, & Forssberg, 2005; Ullen, 2007). Also, in light of evidence that the left supramarginal gyrus is engaged during the manipulation of sequences of phonemes in working memory, the signal increases in this region may reflect the development of phonological representations for the "words" in the artificial languages (Gelfand & Bookheimer, 2003; Ravissa, Delgado, Chein, Becker, & Fiez, 2004; Xiao et al., 2005). While children's signal increases in superior temporal and inferior parietal cortices indicate that they were implicitly segmenting the speech streams, the fact that these increases were bilateral suggests that the calculation of the statistical regularities might be more difficult for children, based on prior work showing that right hemisphere homologues of canonical language areas in the left hemisphere tend to be recruited in linguistic tasks as a function of difficulty or complexity (Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Hasegawa, Carpenter, Just, & 2002; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Lee & Dapretto, 2006). Also, increased lateralization to the left hemisphere has been observed in event-related potential (ERP) studies conducted in young children in response to known words, whereas a more bilateral ERP response has been found for words that are not yet in their vocabulary (Mills, Coffey-Corina, & Neville, 1993, 1997).

Perhaps most interestingly, unlike adults, who displayed signal increases only during exposure to the Stressed and Unstressed Language streams, significant signal increases over time in bilateral superior temporal gyri were also observed when children listened to the Random Syllables stream. This suggests that perhaps children are more sensitive to the very small statistical regularities that are present in that stream despite the fact that it could not be as readily segmented as the artificial language streams (Figure 4 and Table 4). Given that the input consisted of the repetition of just twelve syllables, many pairs of syllables co-occurred within the Random syllables stream (between 1–10 times), hence providing some statistical regularities that, albeit very small, could lead to the parsing of the stream into many different two-syllable words. These signal increases for the Random syllables stream, coupled with the bilateral pattern of increases for the artificial language streams, indicate that, as children

calculate statistical regularities between syllables, they may also begin to map possible two-syllable words.

In addition to the signal increases over time in superior temporal and inferior parietal regions, children also displayed signal increases in bilateral putamen and the left dorsal premotor/supplementary motor area while listening to the artificial languages (Figure 4 and Table 4). These signal increases over time in the putamen in children lend support to recent accounts that the basal ganglia, in addition to their role in domain-general sequence learning (e.g., Doyon, Penhune, & Ungerleider, 2003; Poldrack & Gabrieli, 2001; Poldrack et al., 2005; Saint-Cyr, 2003; Van der Graaf, de Jong, Maguire, Meiners, & Leenders, 2004), may serve to initiate the cortical representation of phonological information in temporal and inferior frontal regions, thus facilitating the binding of phonemes together into higher order structures such as words (Booth, Wood, Lu, Houk, & Bitan, 2007; Ullman, 2001; Ullman, 2004). Further, learning-related increases in activity in the putamen have been noted during the acquisition of an artificial grammar (Newman-Norlund et al., 2006), as well as during the learning of motor sequences in serial reaction time tasks (Muller, Kleinhans, Pierce, Kemmotsu, & Courchesne, 2002; Seidler et al., 2005; Van der Graaf et al., 2004). While the dorsal premotor and supplementary motor areas have typically been associated with the planning and production of complex sequences of movement, recent studies have demonstrated that this region is also activated during the processing of abstract, movement-independent temporal sequences in conjunction with superior temporal and inferior frontal regions (Ullén et al., 2005; Ullén, 2007). Supplementary motor area activation has also been commonly used as an index of sub-vocal rehearsal during speech perception and is indicative of a role for this region in verbal working memory (for a review, see Baddeley, 2003; Gruber, Kleinschmidt, Binkofski, Steinmetz, & von Cramon, 2000; Smith & Jonides, 1999; Woodward et al., 2006). Further, given increasing evidence of the involvement of the motor system in speech comprehension at levels ranging from phoneme perception to narrative comprehension (Wilson, Saygin, Sereno, & Iacoboni, 2004; Wilson & Iacoboni, 2006; Wilson et al., 2007), one interpretation of the signal increases seen in the SMA is that they may reflect a mapping of the sequential temporal structure of the syllables as the statistical regularities between them are being computed online in temporal regions during exposure to the artificial language streams.

Whereas signal increases during listening to the artificial languages were observed in both children and adults in temporal and parietal regions, adults displayed greater signal increases than children in bilateral superior and middle temporal cortices, and children displayed greater signal increases than adults in left inferior parietal and dorsal premotor/supplementary motor areas (Figure 6 and Table 9). Adults' greater signal increases in temporal regions for the artificial language streams, coupled with their faster reaction times than children on the post-scan Word Discrimination Task, suggest that adults may be relatively more successful at tracking high statistical regularities when attempting to segment these artificial language speech streams. This pattern of the results, while contrary to what might be expected given the documented language-learning advantage for children (e.g., Johnson & Newport, 1989; Weber-Fox & Neville, 2001), could be due in part to the use of native English phonemes and syllables used to form the speech streams and the fact that adults' experience with processing them far exceeds that of children. Adults' greater experience with these stimuli as well as with the pattern of stress typical in the English language might indeed have a significant impact on the neural processing of these cues, particularly in light of evidence that listeners may apply native segmentation procedures when attempting to learn new languages (Cutler 2000–2001; Cutler, Mehler, Norris, & Segui, 1986; Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993; Finn & Kam, 2008; Snijders & Hagoort, 2007; Vroomen, Tuomainen, & de Gelder, 1998). In addition, the fact that adults displayed greater signal increases for the Stressed Language compared to the

Unstressed Language streams in bilateral superior temporal cortices whereas children displayed the opposite pattern suggests that perhaps the more experience a listener has with the pattern of stress falling predominantly on the initial syllable of English words (Cutler & Carter, 1987), the more he or she might rely on those cues as a marker for word boundaries. The notion that it might be more difficult for children to utilize the statistical regularities present in the streams to identify word boundaries is also supported by our finding that children showed greater signal increases than adults while listening to the artificial languages in regions that subserve phonological processing and sub-vocal rehearsal in working memory and the mapping of the sequential temporal structure of syllables (Figure 6 and Table 9).

Correlations with Behavioral Word Discrimination Accuracy and Reaction Time

Overall, children could not reliably identify the words used to create the artificial language streams in the post-scan behavioral test, similar to what was found in our previous study in adults (McNealy et al., 2006). Nevertheless, we propose that the observed significant signal increases over time in temporal, frontal, parietal and subcortical regions during exposure to the artificial languages indicate that initial, implicit word segmentation might have occurred, as it is not uncommon for changes in neural activity to precede a behavioral change in task performance (e.g., Jaeggi et al., 2003; Kelly and Garavan, 2005; Landau, Schumacher, Garavan, Druzgal, & D'Esposito, 2004; McLaughlin et al., 2004; Shadmehr and Holcomb, 1997; Shestakova et al., 2003). Indeed, such early neural changes have been reported in an ERP study of word segmentation and in our previous fMRI study in adults (Sanders et al., 2002; McNealy et al., 2006). Therefore, if the observed signal increases over time during the Speech Stream Exposure Task reflect implicit word segmentation, a positive relationship should be observed between this activity and children's ability to discriminate words from within the artificial language streams. Just as was seen in adults, children displayed a positive relationship between activity within the left STG during exposure to the artificial languages (as compared to the Random Syllables stream) and accuracy at discriminating words in the post-scan behavioral Word Discrimination Task. Further, faster responses to words in the post-scan test were associated with relatively greater activity in bilateral STG, supporting our interpretation that these regions likely play a critical role in enabling initial, implicit word segmentation in children. It should be noted that the ability to discriminate words from partwords and nonwords does not necessarily indicate successful segmentation of the speech stream into word-like units. However, results from a behavioral study in which learning of object labels was facilitated by prior experience with segmenting a continuous speech stream that contained the novel object names as words strongly suggests that learners go on to utilize their knowledge of statistical regularities for actual segmentation and identification of words (Graf Estes et al., 2007).

Word Discrimination Task and Developmental Differences in Neural Activation

Further evidence that suggesting that word segmentation might have implicitly occurred when children listened to the artificial language streams during the Speech Stream Exposure Task is provided by the results of the Word Discrimination Task, where participants were exposed to trisyllabic combinations that had occurred with different frequencies within the speech streams. As was observed in adults, children displayed greater activity in frontal regions in response to trisyllabic combinations with higher frequency of occurrence within the speech streams (i.e., higher transitional probabilities between syllables). In particular, listening to Words as compared to Nonwords (i.e., trisyllabic combinations that had occurred 45 times within the artificial language streams vs. 1 time within the random syllables stream) elicited greater activity in left IFG and right CG (Figure 7). Given that the left IFG is consistently engaged by various tasks involving phonological and sequential processing (see Bookheimer, 2002, for a review), such as phonemic discrimination, temporal

sequencing, articulatory recoding, and the maintenance of acoustic information in working memory (e.g., Burton, Small, & Blumstein, 2000; Newman & Twieg, 2001; Gelfand & Bookheimer, 2003; LoCasto, Krebs-Noble, Gullapalli, & Burton, 2004; Xiao et al., 2005), we previously proposed that greater IFG activity in response to words compared to nonwords may reflect the initial phonological representation for words in the artificial languages (McNealy et al., 2006).

The finding that the left IFG was also significantly active in the comparison of Partwords (i.e., trisyllabic combinations that had occurred 15 times within the artificial language streams) versus Nonwords lends further support to the notion that implicit word segmentation might have taken place during exposure to the speech streams, particularly considering that all of the stimuli (Words, Partwords, and Nonwords) are pseudowords that would otherwise not be processed differently from one another. These differences in neural activity in response to Words, Partwords and Nonwords could simply be due to the different frequencies with which they occurred in the speech streams, but may also represent the outcome of having calculated the transitional probabilities between the syllables of these items. This latter interpretation is supported by several behavioral studies in which it was shown that listeners are sensitive to transitional probabilities above and beyond frequency of co-occurrence of adjacent syllables (Aslin et al., 1998) and that the strength of transitional probabilities between syllables is predictive of behavioral responses (Saffran, Newport et al., 1996).

Unlike adults, there were no significant differences in activity for the Words compared to Partwords in children. This lack of differential activity might be due to children having partial recognition of the Partwords as potential words after hearing them fifteen times in the artificial language streams. Alternatively, it might be due to children having not implicitly learned the Words well enough to distinguish them from Partwords. The fact that adults displayed greater activity than children in left inferior and middle frontal gyri for Words compared to both Nonwords and Partwords is in line with the interpretation of the results from the Speech Stream Exposure Task that children may not be as successful as adults at discriminating the words within the artificial language streams (Figure 8; Table 10). In addition, our finding that adults, and not children, exhibited greater activity when listening to words and partwords from the Stressed Language as compared to words and partwords from the Unstressed Language might indicate that children do not benefit as much as adults do from the presence of prosodic cues. Finally, the anterior cingulate has been shown to subserve a variety of functions related to conflict monitoring and top-down control of cognitive processing (Barch et al., 2001; Kerns, Cohen, Stenger, & Carter, 2004; Kerns et al., 2004; van Veen & Carter, 2002), and the greater involvement of this region in children than adults when they listened to Words compared to Nonwords may reflect control processes engaged as children attempt to retrieve initial word representations (Figure 8).

Conclusion and Future Directions

Together with previous behavioral work (e.g., Graf Estes et al., 2007) and our prior neuroimaging findings in adults, the current study in children suggests that minimal exposure to a continuous stream of speech containing statistical and prosodic cues may be sufficient for implicit word segmentation to occur and provides a neural signature of online, implicit speech parsing in the developing brain. While the majority of neuroimaging studies of language have highlighted the large degree of similarity between children and adults in the networks that support linguistic processing once a language is acquired (for a review, see Friederici, 2006), the current research shows that the actual learning process itself is not subserved by identical neural mechanisms across development and offers an initial characterization of developmental differences in the neural basis of speech parsing. At an applied level, this paradigm may be useful for identifying abnormalities in the neural

architecture underlying language learning in developmental disorders impacting language acquisition, as well as for exploring changes in this circuitry following interventions.

These findings have theoretical implications as well. It has recently been proposed (Kuhl, 2004) that a continuous process of neural commitment to the statistical and prosodic patterns of the language experienced early in life facilitates future learning of patterns that match those already learned and inhibits the learning of those that do not match. Understanding how such a process occurs could help to account for the well-documented decrease in the ability to acquire another language with native-like proficiency after puberty (e.g., Johnson & Newport, 1989; Weber-Fox & Neville, 2001). Although the neural commitment hypothesis has been outlined in more detail for phonetic learning during first language acquisition in infancy (see Werker & Tees, 2005 for a review), it has been posited that the same basic principles may be applicable to other aspects of language learning such as word segmentation (Kuhl, 2004). In this context, the current study may lend insight into how the neural commitment process might be instantiated in the developing brain.

Our findings demonstrate the existence of developmental changes in the neural mechanisms by which cues to word boundaries are processed, with adults engaging a more focal network comprised of canonical language cortices and children recruiting a wider network including regions involved in attention and working memory. The fact that children exhibited greater recruitment of these dorsal parietal networks while listening to all the speech streams and that they displayed signal increases in superior temporal cortices as a function of exposure to the random syllables stream supports the notion that children may perhaps be more sensitive to small transitional probabilities than adults are. At the same time, our findings suggest that adults might be more successful than children at segmenting the current artificial language streams because adults: i) displayed greater signal increases than children in temporal cortices, where activity correlated with better identification of words in the artificial languages, ii) were faster than children to respond when identifying words on the behavioral post-test, and iii) exhibited greater differential activity between words and nonwords in the Word Discrimination Task. Despite the well-recognized advantage of children for learning new languages, an increased commitment of language-relevant cortex to the parsing of the artificial language streams with age might thus confer an advantage to adults. However, when exposed to a *real* novel language, where transitional probabilities between syllables are considerably lower given the large corpus of new words, it is possible that the ability to detect subtle statistical regularities - perhaps afforded by greater engagement of working memory and attentional networks - could be advantageous. This may be particularly so for languages that differ more substantially from one's native language. Future fMRI studies of word segmentation using sequences comprised of non-native phonemes or nonlinguistic stimuli such as tones to ensure that prior experience would be the same for children and adults, are needed in order to directly address whether developmental changes in the mechanism of speech parsing might play a role in the age-related decline in the ability to learn new languages with native-like proficiency.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Santa Fe Institute Consortium, the National Science Foundation, and the FPR-Center for Culture, Brain and Development. For generous support the authors also wish to thank the Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, The Ahmanson Foundation, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson

Family and Northstar Fund. The project described was supported by Grant Numbers RR12169, RR13642 and RR00865 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH); its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCR or NIH. Thanks also to Larissa Borofsky, Nicole Vazquez and Jennifer Pfeifer for their help with data collection and analysis.

References

- Achenbach, TM.; Edelbrock, CS. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: Department of Psychiatry, University of Vermont; 1983.
- Aslin RN, Saffran JR, Newport EL. Computation of probability statistics by 8-month-old infants. *Psychological Science*. 1998; 9:321–324.
- Baddeley A. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*. 2003; 4:829–839.
- Barch DM, Braver TS, Akbudak E, Conturo T, Ollinger J, Snyder A. Anterior cingulate cortex and response conflict: Effects of response modality and processing domain. *Cerebral Cortex*. 2001; 11:837–848. [PubMed: 11532889]
- Bischoff-Grethe A, Proper SM, Mao H, Daniels KA, Berns GS. Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *The Journal of Neuroscience*. 2000; 20(5):1975–1981. [PubMed: 10684898]
- Blakemore SJ, Rees G, Frith CD. How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia*. 1998; 36(6):521–529. [PubMed: 9705062]
- Bookheimer S. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Reviews in Neuroscience*. 2002; 25:151–188.
- Booth JR, Wood L, Lu D, Houk J, Bitan T. The role of the basal ganglia and cerebellum in language processing. *Brain Research*. 2007; 1133:136–144. [PubMed: 17189619]
- Booth JR, Cho S, Burman DD, Bitan T. Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. *Developmental Science*. 2007; 10(4):441–451. [PubMed: 17552934]
- Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL. Developmental changes in human cerebral functional organization for word generation. *Cerebral Cortex*. 2005; 15:275–290. [PubMed: 15297366]
- Burton MW, Small SL, Blumstein SE. The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*. 2000; 12(4):679–690. [PubMed: 10936919]
- Callan AM, Callan DE, Masaki S. When meaningless symbols become letters: Neural activity change in learning new phonograms. *NeuroImage*. 2005; 1–10. doi:03296.
- Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*. 2003; 19:113–124. [PubMed: 12781731]
- Carpenter PA, Just MA, Keller RA, Eddy WF, Thulborn KR. *NeuroImage*. 1999; 10:216–224. [PubMed: 10417254]
- Casey BJ, Galvan A, Hare TA. Changes in cerebral functional organization during cognitive development. *Current Opinion in Neurobiology*. 2005; 15:239–244. [PubMed: 15831409]
- Casey BJ, Giedd JN, Thomas KM. Structural and functional brain development and its relation to cognitive development. *Biological Psychology*. 2000; 54:241–257. [PubMed: 11035225]
- Cavanna AE, Trimble MR. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*. 2006; 129:564–583. [PubMed: 16399806]
- Chou TL, Booth JR, Burman DD, Bitan T, Bigio JD, Lu D, Cone NE. Developmental changes in the neural correlates of semantic processing. *NeuroImage*. 2006; 29:1141–1149. [PubMed: 16275017]
- Cole, R.; Jakimik, J. A model of speech perception. In: Cole, R., editor. *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum; 1980. p. 133–163.
- Culham JC, Kanwisher NG. Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*. 2001; 11:157–163. [PubMed: 11301234]
- Cutler A. Listening to a second language through the ears of a first. *Interpreting*. 2000–2001; 5:1–23.

- Cutler A, Carter DM. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*. 1987; 2:133–142.
- Cutler A, Mehler J, Norris D, Segui J. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*. 1986; 25:385–400.
- Cutler A, Otake T. Mora or phoneme – further evidence for language-specific listening. *Journal of Memory and Language*. 1994; 33:824–844.
- Doyon J, Penhune V, Ungerleider LG. Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*. 2003; 41:252–262. [PubMed: 12457751]
- Durston S, Casey BJ. What have we learned about cognitive development from neuroimaging? *Neuropsychologia*. 200510.1016/j.neuropsychologia.2005.10.010
- Durston S, Davidson MC, Tottenham N, Galvan A, Spicer J, Fossella J, Casey BJ. A shift from diffuse to focal cortical activity with development. *Developmental Science*. 2006; 9(1):1–20. [PubMed: 16445387]
- Finn AS, Kam CLH. The curse of knowledge: First language knowledge impairs adults learners' use of novel statistics for word segmentation. *Cognition*. 200810.1016/j.cognition.2008.04.002
- Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*. 2002; 99(24):15822–15826.
- Fletcher PC, Zafiris O, Frith CD, Honey RAE, Corlett PR, Zilles K, Fink GR. On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cerebral Cortex*. 2004:1–14. doi10.1093.
- Friederici AD. The neural basis of language development and its impairment. *Neuron*. 2006; 52:941–952. [PubMed: 17178399]
- Friederici AD, Steinhauer K, Pfeifer E. Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*. 2002; 99(1):529–534.
- Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ. Multisubject fMRI studies and conjunction analyses. *NeuroImage*. 1999; 10:385–396. [PubMed: 10493897]
- Gathercole SE. Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*. 1999; 3:410–419. [PubMed: 10529796]
- Gelfand J, Bookheimer SY. Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron*. 2003; 38:831–842. [PubMed: 12797966]
- Golestani N, Zatorre RJ. Learning new sounds of speech: reallocation of neural substrates. *NeuroImage*. 2004; 21:494–506. [PubMed: 14980552]
- Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*. 2007; 18(3):254–260. [PubMed: 17444923]
- Gruber O, Kleinschmidt A, Binkofski F, Steinmetz H, von Cramon DY. Cerebral correlates of working memory for temporal information. *Neuroreport*. 2000; 11(8):1689–1693. [PubMed: 10852226]
- Hasegawa M, Carpenter PA, Just MA. An fMRI study of bilingual sentence comprehension and workload. *NeuroImage*. 2002; 15:647–660. [PubMed: 11848708]
- Hashimoto R, Sakai KL. Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*. 2004; 42:311–322. [PubMed: 15091345]
- Hauser MD, Newport EL, Aslin RN. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*. 2001; 78:B53–B64. [PubMed: 11124355]
- Hickok G, Poeppel D. The cortical organization of speech processing. *Nature Reviews Neuroscience*. 2007; 8:393–402.
- Holland SK, Plante E, Weber Byars A, Strawsburg RH, Schmithorst VJ, Ball WS Jr. Normal fMRI brain activation patterns in children performing a verb generation task. *NeuroImage*. 2001; 14(4): 837–43. [PubMed: 11554802]
- Jaeggi SM, Seewer R, Nirkko AC, Eckstein D, Schroth G, Groner R, Gutbrod K. Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single

- and dual tasks: functional magnetic resonance imaging study. *NeuroImage*. 2003; 19:210–225. [PubMed: 12814572]
- Johnson EK, Newport EL. Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*. 1989; 21:60–99. [PubMed: 2920538]
- Johnson EK, Jusczyk PW. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language*. 2001; 44:548–567.
- Jusczyk PW. How infants adapt speech-processing capacities to native-language structure. *Current Directions in Psychological Science*. 2002; 11(1):15–18.
- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. Brain activation modulated by sentence comprehension. *Science*. 1996; 274:114–116. [PubMed: 8810246]
- Kelly AMC, Garavan H. Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*. 2005; 15:1089–1102. [PubMed: 15616134]
- Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. *Science*. 2004; 303:1023–1026. [PubMed: 14963333]
- Kerns JG, Cohen JD, Stenger VA, Carter CS. Prefrontal cortex guides context-appropriate responding during language production. *Neuron*. 2004; 43:283–291. [PubMed: 15260963]
- Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*. 2002; 83:B35–42. [PubMed: 11869728]
- Kuhl PK. Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*. 2004; 5:831–843.
- Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*. 2006; 9(2):F13–F21. [PubMed: 16472309]
- Landau SM, Schumacher EH, Garavan H, Druzgal TJ, D'Esposito M. A functional MRI study of the influence of practice on component processes of working memory. *NeuroImage*. 2004; 22:211–221. [PubMed: 15110011]
- Lee SS, Dapretto MD. Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *NeuroImage*. 2006; 29:536–544. [PubMed: 16165371]
- LoCasto PC, Krebs-Noble D, Gullapalli RP, Burton MW. An fMRI investigation of speech and tone segmentation. *Journal of Cognitive Neuroscience*. 2004; 16(9):1612–1624. [PubMed: 15601523]
- Luna B, Garver KE, Urban TA, Lazar NA, Sweeney JA. Maturation of cognitive processes from late childhood to adulthood. *Child Development*. 2004; 75:1357–1372. [PubMed: 15369519]
- Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*. 2002; 82:B101–B111. [PubMed: 11747867]
- McLaughlin J, Osterhout L, Kim A. Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*. 2004; 7(7):703–704.
- McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*. 2006; 26(29):7629–7639. [PubMed: 16855090]
- Mills DL, Coffey-Corina S, Neville HJ. Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*. 1993; 5:326–342.
- Mills DL, Coffey-Corina S, Neville HJ. Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*. 1997; 13:397–445.
- Müller RA, Kleinhans N, Pierce K, Kemmotsu N. Functional MRI of motor sequence acquisition: effects of learning stage and performance. *Cognitive Brain Research*. 2002; 14:277–293. [PubMed: 12067701]
- Mutti, M.; Sterling, HM.; Martin, NA.; Spalding, NV. *Quick Neurological Screening Test*. 2. Novato, CA: Academic Therapy Publications; 1998. revised edition
- Newman-Norlund RD, Frey SH, Petitto L-A, Grafton ST. Anatomical substrates of visual and auditory miniature second-language learning. *Journal of Cognitive Neuroscience*. 2006; 18(12):1984–1997. [PubMed: 17129186]

- Newman R, Ratner NB, Jusczyk AM, Jusczyk PW, Dow KA. Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*. 2006; 42(4):643–655. [PubMed: 16802897]
- Newman SD, Twieg D. Differences in auditory processing of words and pseudowords: An fMRI study. *Human Brain Mapping*. 2001; 14:39–47. [PubMed: 11500989]
- Noppeney U, Price CJ. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience*. 2004; 16(4):702–713. [PubMed: 15165357]
- Olesen PJ, Westerberg H, Klingberg T. Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*. 2004; 7:75–79.
- Opitz B, Friederici AD. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage*. 2003; 19:1730–1737. [PubMed: 12948727]
- Opitz B, Friederici AD. Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *The Journal of Neuroscience*. 2004; 24(39):8436–8440. [PubMed: 15456816]
- Osaka M, Komori M, Morishita M, Osaka N. Neural bases of focusing attention in working memory: An fMRI study based on group differences. *Cognitive, Affective, and Behavioral Neuroscience*. 2007; 7(2):130–139.
- Otake T, Hatano G, Cutler A, Mehler J. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*. 1993; 32:258–278.
- Plante E, Holland SK, Schmithorst VJ. Prosodic processing by children: an fMRI study. *Brain and Language*. 2006; 97(3):332–342. [PubMed: 16460792]
- Poldrack RA, Gabrieli JDE. Characterizing the neural mechanisms of skill learning and repetition priming. *Brain*. 2001; 124:67–82. [PubMed: 11133788]
- Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, Knowlton BJ. The neural correlates of motor skill automaticity. *The Journal of Neuroscience*. 2005; 25(22):5356–5364. [PubMed: 15930384]
- Rauschecker AM, Pringle A, Watkins KE. Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*. 2007; 28(10):1002–1010. [PubMed: 17444660]
- Ravissa SM, Delgado MR, Chein JM, Becker JT, Fiez JA. Functional dissociations within the inferior parietal cortex in verbal working memory. *NeuroImage*. 2004; 22:562–573. [PubMed: 15193584]
- Robertson EM. The serial reaction time task: Implicit motor skill learning? *The Journal of Neuroscience*. 2007; 27(38):10073–10075. [PubMed: 17881512]
- Saint-Cyr JA. Frontal-striatal circuit functions: Context, sequence, and consequence. *Journal of the International Neuropsychological Society*. 2003; 9:103–128. [PubMed: 12570364]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. [PubMed: 8943209]
- Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. *Cognition*. 1999; 70:27–52. [PubMed: 10193055]
- Saffran JR, Newport EL, Aslin RN. Word segmentation: the role of distributional cues. *Journal of Memory and Language*. 1996; 35:606–621.
- Saffran JR, Wilson DP. From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy*. 2003; 4(2):273–284.
- Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*. 2002; 5(7):700–703.
- Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE. Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science*. 2002; 296:1476–1479. [PubMed: 12029136]
- Schmithorst VJ, Holland SK, Plante E. Cognitive modules utilized for narrative comprehension in children: A functional magnetic resonance imaging study. *NeuroImage*. 2006; 29:254–266. [PubMed: 16109491]

- Seidler RD, Purushotham A, Kim SG, Ugurbil K, Willingham D, Ashe J. Neural correlates of encoding and expression in implicit sequence learning. *Experimental Brain Research*. 2005; 165:114–124.
- Shadmehr R, Holcomb HH. Neural correlates of motor memory consolidation. *Science*. 1997; 277:821–825. [PubMed: 9242612]
- Shestakova A, Huotilainen M, Ceponiene R, Cheour M. Event-related potentials associated with second language learning in children. *Clinical Neurophysiology*. 2003; 114:1507–1512. [PubMed: 12888034]
- Smith EE, Jonides J. Storage and executive processes in the frontal lobes. *Science*. 1999; 283:1657–1661. [PubMed: 10073923]
- Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*. 200710.1016/j.cognition.2007.06.010
- Snijders TM, Kooijman V, Cutler A, Haggort P. Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*. 2007; 1178:106–113. [PubMed: 17931604]
- Thiel CM, Shanks DR, Henson RNA, Dolan RJ. Neuronal correlates of familiarity-driven decisions in artificial grammar learning. *Neuroreport*. 2003; 14:131–136. [PubMed: 12544844]
- Thiessen ED, Saffran JR. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9- month-old infants. *Developmental Psychology*. 2003; 39(4):706–716. [PubMed: 12859124]
- Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF. Development of neural mechanisms for reading. *Nature Neuroscience*. 2003; 6:767–773.
- Ullen F. Independent neural control of rhythmic sequences – behavioral and fMRI evidence. *Physiology and Behavior*. 2007; 92:193–198. [PubMed: 17568634]
- Ullen F, Bengtsson SL, Ehrsson HH, Forssberg H. Neural control of rhythmic sequences. *Annals of the New York Academy of Sciences*. 2005; 1060:368–376. [PubMed: 16597788]
- Ullman MT. A neurocognitive perspective on language: The declarative/procedural model. *Nature Neuroscience Reviews*. 2001; 2:717–726.
- Ullman MT. Contributions of memory circuits to language: the declarative/procedural model. *Cognition*. 2004; 92:231–270. [PubMed: 15037131]
- Van der Graaf FHCE, de Jong BM, Maguire RP, Meiners LC, Leenders KL. Cerebral activation related to skills practice in a double serial reaction time task: Striatal involvement in random-order sequence learning. *Cognitive Brain Research*. 2004; 20:120–131. [PubMed: 15183385]
- Van Veen V, Carter CS. The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology and Behavior*. 2002; 77:477–482. [PubMed: 12526986]
- Vroomen J, Tuomainen J, de Gelder B. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*. 1998; 38:133–149.
- Weber-Fox C, Neville HJ. Sensitive periods differentiate processing of open- and closed-class words: an ERP study of bilinguals. *Journal of Speech, Language and Hearing Research*. 2001; 44(6): 1338–1353.
- Werker JF, Tees RC. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*. 2005; 46:233–251. [PubMed: 15772961]
- Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*. 2006; 33:316–325. [PubMed: 16919478]
- Wilson SM, Molnar-Szakacs I, Iacoboni M. Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex*. 200710.1093/cercor/bhm049
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*. 2004; 7(7):701–702.
- Wonnacott E, Newport EL, Tanenhaus MK. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*. 200710.1016/j.cogpsych.2007.04.002

- Woods RP, Dapretto M, Sicotte NL, Toga AW, Mazziotta JC. Creation and use of a Talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. *Human Brain Mapping*. 1999; 8(2–3):73–79. [PubMed: 10524595]
- Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*. 1998; 22:139–152. [PubMed: 9448779]
- Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC. Automated image registration: II. Intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography*. 1998; 22:153–165. [PubMed: 9448780]
- Woodward TS, Cairo TA, Ruff CC, Takane Y, Hunter MA, Ngan ETC. Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience*. 2006; 139:317–325. [PubMed: 16324799]
- Xiao Z, Zhang JX, Wang X, Wu R, Hu X, Weng X, Tan LH. Differential activity in left inferior frontal gyrus for pseudowords and real words: An event-related fMRI study on auditory lexical decision. *Human Brain Mapping*. 2005; 25:212–221. [PubMed: 15846769]

Figure 1.

In the Speech Stream Exposure Task, three sets of twelve 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 pseudo-randomly concatenating individual syllables (B). Participants listened to three counterbalanced speech streams (C) containing strong statistical cues (Unstressed Language), strong statistical and prosodic cues (Stressed Language), and no reliable cues to guide word segmentation (Random Syllables).

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 previously heard by a participant (B).

Figure 3.

Extensive activity in temporal and parietal cortices was observed when children listened to the Random Syllables stream (A), the Unstressed Language stream (B), and the Stressed Language stream (C) compared to resting baseline. For display purposes only, activation maps are thresholded at $p < 0.05$ for both magnitude and spatial extent, corrected for multiple comparisons at the cluster level.

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↑; A) as well as for the Random Syllables condition (R↑; B), thresholded at $p < 0.05$ for both magnitude and spatial extent, corrected for multiple comparisons at the cluster level (except for the LH of R↑).

Figure 5. While listening to the Unstressed and Stressed Language streams (U + S) during the Speech Stream Exposure Task, children displayed greater activity than adults in dorsal parietal regions (A), whereas adults displayed greater activity than children in bilateral temporal and inferior frontal cortices (B), thresholded at $p < 0.05$ for both magnitude and spatial extent, corrected for multiple comparisons at the cluster level.

Figure 6.

Children displayed reliably greater signal increases than adults over the course of listening to the Stressed and Unstressed Language streams ($U\uparrow+S\uparrow$) in left IPL and SMA (A), and adults displayed reliably greater signal increases than children in bilateral STG (B). Children also displayed reliably greater signal increases than adults over the course of listening to the Random Syllables stream ($R\uparrow$) in bilateral STG and left IPL (C), thresholded at $p < 0.05$ for both magnitude and spatial extent, corrected for multiple comparisons at the cluster level.

Figure 7. Reliable activity was observed within the left inferior frontal gyrus and bilateral anterior cingulate gyrus as participants listened to Words as compared to Nonwords (W-NW) during the Word Discrimination Task, thresholded at $p < 0.01$ for magnitude and $p < 0.05$ for spatial extent, corrected for multiple comparisons at the cluster level.

Figure 8. When listening to Words as compared to Nonwords (W-NW) during the Word Discrimination Task, children displayed significantly greater activity than adults in the right cingulate gyrus (A), and adults displayed significantly greater activity than children in left inferior and middle frontal gyri (B) during listening to words compared to nonwords, thresholded at $p < 0.01$ for magnitude and $p < 0.05$ for spatial extent, corrected for multiple comparisons at the cluster level.

Table 1

Behavioral Performance on the Word Discrimination Task

	Accuracy (% correct)		Response Time (seconds)	
	Words	Partwords	Words	Partwords
Unstressed Language	48.9 (24.8)	41.0 (23.2)	1.859 (0.446)	1.955 (0.460)
Stressed Language	51.3 (23.0)	49.0 (26.9)	1.774 (0.426)	2.035 (0.494)
Random Syllables	56.7 (24.9)	45.6 (26.2)	1.839 (0.427)	1.972 (0.419)

Values presented as mean (SD).

Table 2

Peaks of Activation during Exposure to the Speech Streams

Anatomical Regions	All Activation > Rest					
	BA	x	y	z	t	
Transverse Temporal Gyrus	41 R	42	-22	10	15.69	
Superior Temporal Gyrus	22 L	-48	-30	10	12.82	
	22 R	58	-16	4	14.60	
	42 L	-60	-22	6	15.00	
	42 R	48	-16	6	15.38	
Middle Temporal Gyrus	21 L	-64	-34	-6	4.80	
Inferior Parietal Lobule	40 L	-42	-42	52	7.82	
	40 R	28	-40	56	7.27	
Superior Parietal Lobule	7 L	-34	-46	60	6.98	
Precuneus	7 L	-18	-50	64	6.98	
	7 R	8	-58	62	7.93	
Paracentral Lobule	5 L	-14	-32	62	4.85	
	5 R	10	-32	62	6.23	
Inferior Frontal Gyrus	44 L	-54	4	26	5.59	
Dorsal Premotor Cortex/Supplementary Motor Area	6 L	-28	2	52	6.05	
	6 R	18	-10	52	6.04	
Middle Frontal Gyrus	10 L	-32	42	-10	3.46	
Precentral Gyrus	4 L	-48	-14	48	4.17	
	4 R	24	-24	54	5.07	
	6 L	-54	-10	38	6.27	
	6 R	50	-4	36	6.58	
Postcentral Gyrus	3 L	-54	-18	40	5.59	
	3 R	36	-28	48	6.85	
Anterior Cingulate Gyrus	32 L	-14	24	42	3.78	
Medial Frontal Gyrus	8 L	-12	26	46	3.80	
	10 L	-22	42	-10	3.70	
Superior Frontal Gyrus	9 L	-20	44	36	3.16	

All Activation > Rest					
Anatomical Regions	BA	x	y	z	t
Insula	10	L -8	54	0	3.48
		L -38	-8	12	3.20

Activity thresholded at $t > 3.13$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$). BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 3
Peaks where Activity Differed during Exposure to the Random Syllables Stream and the Artificial Languages

Anatomical Regions	BA	Random Syllables > Unstressed Language				Random Syllables > Stressed Language				Unstressed Language > Random Syllables				Stressed Language > Random Syllables			
		x	y	z	t	x	y	z	t	x	y	z	t	x	y	z	t
Precuneus	7 R	20	-38	46	3.67												
Dorsal Premotor Cortex/ Supplementary Motor Area	6 L	-18	10	52	3.47	-32	-8	50	3.39								
	6 R	32	-8	54	3.76												
Precentral Gyrus	6 L	-46	-6	22	3.58												
	6 R					42	-4	30	3.44								
Insula	L	38	-16	10	3.44												
Superior Temporal Gyrus	22 L									-62	-30	6	3.27				
	22 R													58	-40	6	3.61
Inferior Parietal Lobule	40 L									-48	-48	28	3.31				
Basal Ganglia	L									-16	0	4	3.29*				

Activity thresholded at $t > 3.13$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$).

* indicates a small volume correction was applied. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 4
Peaks of Activation for Regions where Activity Increased as a Function of Exposure to the Speech Streams

Anatomical Regions		Increases in Artificial Languages (U↑ + S↑)								Increases in Random Syllables (R↑)								Increases in Artificial Languages > Random Syllables (U↑ + S↑ > R↑)			
		BA	x	y	z	t	x	y	z	t	x	y	z	t	x	y	z	t			
Superior Temporal Gyrus	22 L	-56	-12	0	3.85	-52	-20	0	4.66*	-48	-12	2	2.77*								
	22 R	50	-36	16	3.74	60	-22	2	3.87												
	42 L	-56	-26	6	3.98	-66	-30	12	2.98												
	42 R	58	-26	8	3.64																
Transverse Temporal Gyrus	41 L	-34	-34	10	2.53																
	41 R	34	-28	14	3.71																
Inferior Parietal Lobule	40 L	-54	-38	36	3.45					-54	-40	34	3.74								
Dorsal Premotor Cortex/ Supplementary Motor Area	6 L	-20	6	50	3.53																
Basal Ganglia	L	-14	8	0	3.10*																
	R	20	10	8	2.81*																

Activity thresholded at $t > 2.4$ ($p < 0.01$), corrected for multiple comparisons at the cluster level ($p < 0.05$).

* indicates a small volume correction was applied. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 5

Peaks of Activation for Regions where Activity Increased as a Function of Exposure to the Unstressed and Stressed Language Speech Streams Compared to the Random Syllables Stream

Anatomical Regions	BA	Increases in Unstressed Language > Random Syllables ($U \uparrow > R \uparrow$)				Increases in Stressed Language > Random Syllables ($S \uparrow > R \uparrow$)				Increases in Unstressed Language > Stressed Language ($U \uparrow > S \uparrow$)			
		x	y	z	t	x	y	z	t	x	y	z	t
Superior Temporal Gyrus	22 L					-56	-10	2	2.65	-46	-34	10	2.98
	42 L	-40	-22	8	2.68								
	42 R									56	-24	6	3.64*
Inferior Parietal Lobule	40 L	-50	-40	34	2.73	-52	-40	36	3.39*				

Activity thresholded at $t > 2.4$ ($p < 0.01$), corrected for multiple comparisons at the cluster level ($p < 0.05$).

* indicates a small volume correction was applied. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 6
Peaks of Activation Where Children Displayed Greater Activity than Adults During Exposure to the Speech Streams

Anatomical Regions	BA	Stressed Language > Rest					Unstressed Language > Rest					Random Syllables > Rest				
		x	y	z	t		x	y	z	t		x	y	z	t	
Inferior Parietal Lobule	40 L						-44	-48	36	3.95		-32	-48	56	3.46	
Superior Parietal Lobule	7 L	-12	-68	54	3.80		-22	-66	52	3.25		-26	-46	66	4.40	
Precuneus	7 L	-12	-58	60	3.53		-12	-46	64	3.91		-22	-60	52	3.28	
	7 R	10	-62	58	5.15		18	-48	64	3.48		12	-46	66	4.14	
Paracentral Lobule	5 R	4	-24	48	3.64							14	-28	62	3.14	
Dorsal Premotor Cortex/Supplementary Motor Area	6 L	-24	-8	62	3.63											
	6 R	26	-8	60	4.89							18	-10	52	3.82	
Basal Ganglia	L	-22	8	-2	3.34*											

Activity thresholded at $t > 3.10$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$).

* indicates a small volume correction was applied. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Peaks of Activation Where Adults Displayed Greater Activity than Children During Exposure to the Speech Streams

Table 7

Anatomical Regions		BA	Stressed Language > Rest						Unstressed Language > Rest						Random Syllables > Rest					
			x	y	z	t	x	y	z	t	x	y	z	t	x	y	z	t		
Inferior Frontal Gyrus	44	R																		
	45	L	-54	22	18	3.19*														
	45	R	46	22	10	3.55	44	28	2	3.11										
	47	R					38	28	2	3.87	36	22	-6	3.41*						
Middle Frontal Gyrus	10	L	-36	48	6	3.72														
	10	R					34	42	-6	3.75										
Medial Frontal Gyrus	10	L	-4	52	12	3.59														
Anterior Cingulate Gyrus	24	L																		
	32	R																		
Precentral Gyrus	6	R					40	-4	38	4.56										
Supplementary Motor Area	6	L	-4	-6	64	3.47														
Superior Temporal Gyrus	22	L	-54	-40	12	4.73	-60	-22	6	4.32	-46	-40	14	5.51						
	22	R	52	-30	16	4.44	54	-6	2	4.67	54	-50	10	4.09						
	42	L	-60	-22	6	4.43	-66	-22	6	3.94	-60	-30	8	4.52						
	42	R					58	-20	8	4.22	56	-28	14	4.40						
Middle Temporal Gyrus	21	L					-64	-40	2	3.88	-64	-44	4	3.62						
	21	R					58	-18	-2	3.41	56	-12	-12	4.98						
Inferior Parietal Lobule	40	L									-52	-46	26	4.68						
Hippocampal Gyrus		R									0	-36	-4	4.40						
Insula		L									-34	-18	0	3.68						
		R									30	-28	20	3.78						
Putamen		L									-28	10	6	3.35						

Activity thresholded at $t > 3.10$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$).

* indicates a small volume correction was applied. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Peaks of Activation where Adults and Children Displayed Differential Activity when Listening to the Random Syllables Stream Compared to the Artificial Languages

Table 8

Random Syllables > Stressed and Unstressed Languages (R > U + S)						
Anatomical Regions	BA	x	y	z	t	
Inferior Frontal Gyrus	44 L	-44	14	10	3.35	
Superior Temporal Gyrus	22 L	-44	-42	16	3.79	
Middle Temporal Gyrus	21 L	-60	-58	4	3.97	
	21 R	54	-12	-12	4.45	
Inferior Parietal Lobule	40 L	-54	-48	32	4.87	
	40 R	50	-34	28	3.92	
	39 L	-50	-66	32	3.13	
Paracentral Lobule	5 L	-2	-26	44	3.54	
	5 R	6	-28	50	3.42	
Insula	R	28	-28	20	3.98	
Putamen	L	-22	-2	10	3.66	
	R	30	-22	-2	3.80	

Activity thresholded at $t > 3.10$ ($p < 0.001$), corrected for multiple comparisons at the cluster level ($p < 0.05$). BA refers to putative Brodmann Area. L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 9

Peaks of Activation Where Children and Adults Differed In Signal Increases Over Time as a Function of Exposure to the Speech Streams

Anatomical Regions	BA	Children > Adults						Adults > Children					
		Increases in Artificial Languages (U↑ + S↑)						Increases in Random Syllables Stream (R↑)					
		x	y	z	t	x	y	z	t	x	y	z	t
Superior Temporal Gyrus	22 L					-44	-42	10	2.72	-62	-6	4	3.64
	22 R					60	-22	0	3.13	50	2	-6	2.79
	42 L					-66	-28	12	3.46				
Middle Temporal Gyrus	21 L									-64	-46	6	4.49
	21 R									54	-8	-6	2.76
Inferior Parietal Lobule	40 L	-42	-44	26	2.83	-48	-48	24	2.92				
Dorsal Premotor Cortex/Supplementary Motor Area	6 L	-20	8	48	4.17								

Activity thresholded at $t > 2.37$ ($p < 0.01$), corrected for multiple comparisons at the cluster level ($p < 0.05$). BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.

Table 10
Peaks of Activation where Adults Display Greater Activation than Children in the 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	-46	14	24	3.18*	-54	8	22	3.86			
Middle Frontal Gyrus	9	L	-44	24	30	4.31							
Inferior Parietal Lobule	40	L								-52	-32	36	4.02

Activity thresholded at $t > 3.45$ ($p < 0.001$), small volume correction applied.

* indicates only survives $p < 0.002$. BA refers to putative Brodmann Area; L and R refer to left and right hemispheres; x, y, and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively; t refers to the highest t score within a region.