The neural basis of non-native speech perception in bilingual children

Pilar Archila-Suerte a,*, Jason Zevin b, Aurora Isabel Ramos a, Arturo E. Hernandez a

a University of Houston, USA
b Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, 1300 York Ave. Box 140, NY, NY 10065, USA

A R T I C L E   I N F O

Article history:
Accepted 15 October 2012
Available online 2 November 2012

Keywords:
Speech
Bilingualism
Development
Neuroscience
Learning

A B S T R A C T

The goal of the present study is to reveal how the neural mechanisms underlying non-native speech perception change throughout childhood. In a pre-attentive listening fMRI task, English monolingual and Spanish–English bilingual children – divided into groups of younger (6–8 yrs) and older children (9–10 yrs) – were asked to watch a silent movie while several English syllable combinations played through a pair of headphones. Two additional groups of monolingual and bilingual adults were included in the analyses. Our results show that the neural mechanisms supporting speech perception throughout development differ in monolinguals and bilinguals. While monolinguals recruit perceptual areas (i.e., superior temporal gyrus) in early and late childhood to process native speech, bilinguals recruit perceptual areas (i.e., superior temporal gyrus) in early childhood and higher-order executive areas in late childhood (i.e., bilateral middle frontal gyrus and bilateral inferior parietal lobule, among others) to process non-native speech. The findings support the Perceptual Assimilation Model and the Speech Learning Model and suggest that the neural system processes phonological information differently depending on the stage of L2 speech learning.

Published by Elsevier Inc.

In the current study, we examine the development of non-native speech perception as it is differentially processed by sequential bilingual children between the ages of 6 and 10. About 65% of bilingual children in the US learn Spanish as their native language at home and English as their second language in elementary school (Bedore and Peña, 2008). Despite the benefits of early discrimination abilities on later language development (Chiappe et al., 2001; Kuhl et al., 2005); thus far, only a handful of studies have used neurophysiological methods to investigate speech perception in children (Cunningham et al., 2000; Dehaene-Lambertz and Gliga, 2004). Moreover, most studies have shed light on the changes that occur during infancy and early childhood (Conboy et al., 2005; Garcia-Sierra et al., 2011; Morr et al., 2002; Redcay et al., 2008; Rivera-Gaxiola et al., 2007), but none have investigated the changes that continue to occur in late childhood. To our knowledge, this is the first fMRI study that looks at how sequential bilingual children process non-native speech as their phonological abilities improve over time. Our goal is, therefore, to unveil how the neural mechanisms underlying non-native speech perception change throughout childhood. We tie our findings presented here to recent data with sequential bilingual adults to expand our conclusions about how development of non-native speech perception changes throughout childhood and early adulthood.

Behavioral evidence on the development of speech perception

Monolinguals

It is well established that infants shift from language-general to language-specific perception and that this results in the improved discrimination of native phonemes (Kuhl et al., 1992, 2008; Werker and Curtin, 2005). After this initial entrenchment, however, the mechanisms of speech perception continue to develop throughout childhood and adolescence as phonemic representations become better defined and organized (Huyck and Wright, 2011; Polka et al., 2001; Sundara et al., 2006; Tsao et al., 2006). By adulthood, native phonemic representations have become relatively stabilized (Hazen and Barrett, 2000). Converging behavioral evidence shows that children and adults process native speech differently. It has been proposed that children and adults exploit different strategies to weigh the acoustic parameters in the speech signal. That is, children pay more attention to vocalic formant transitions than adults because these provide dynamic information that enables children to recover the whole syllable (Nittouer, 2007; Nittouer and Lowenstein, 2009; Nittouer et al., 1993). It has also been found that language experience contributes to the improvement of speech perception. For example, monolingual adults more accurately discriminate the difficult English contrast /d/-/ð/ (e.g., dad vs. that) than monolingual children (Sundara et al., 2006) and Filipino-learning infants cannot discriminate the Tagalog contrast /na/-/ŋa/ at 6-months.
but can do so at 12-months of age, thus indicating that experience with the native language is necessary to discriminate contrasts that have reduced acoustic salience (Narayan et al., 2010). Therefore, in learning to discriminate phonemes, monolinguals commit to the sounds of the native language at the end of their first year and continue to restructure these phonemic representations during childhood as perceptual maturation and language experience increases.

**Bilinguals**

Simultaneous bilingual infants narrow their perceptual skills around the same time as their monolingual counterparts (Werker et al., 2009). However, bilingual infants entrench to two languages instead of one, thereby maintaining the ability to discriminate contrasts across languages (Byers-Heinlein et al., 2010). For example, bilingual infants are able to discriminate highly similar phonemes across languages; as in the case of French–English bilingual 10–12 month olds who can distinguish an alveolar /d/ in English from a dental /d/ in French (Sundara et al., 2008) or Spanish–English 8-month olds who can discriminate Spanish /e/ from English /i/ (Sundara and Scutellaro, 2011). As long as the appropriate method is used, Spanish–Catalan bilingual infants can also discriminate the contrasts /e/-/i/ and /e/-/u/ at 8 and 12-months of age (Albareda-Castelló et al., 2011) and French–English bilingual infants can discriminate the contrast /p/-/b/ in English and French around 14-to-20 months of age (Burns et al., 2007). Like monolinguals, bilinguals also continue refining their speech perception throughout childhood. This has been demonstrated in French–English bilingual adults who better discriminate the contrast /d/-/t/ than French–English bilingual children (Sundara et al., 2006). Together, these studies indicate that continued linguistic experience enhances the discrimination of contrasts in both languages. It has been proposed that monolinguals and bilinguals likely undergo different developmental trajectories (Sebastian-Gallés, 2010; Werker et al., 2009) and recent evidence suggests that bilingual infants first tag languages by differentiating their unique rhythms and then attending to the cues that are important in each language through statistical learning (Sundara and Scutellaro, 2011), as opposed to grouping all phonemes from both languages and then sorting out phonemic boundaries. Therefore, despite the timeline similarities between monolinguals and bilinguals, they appear to be tackling the problem of phonemic discrimination in different ways.

**Neurophysiological evidence on the development of speech perception**

**Monolinguals**

Some studies have used neurophysiological measures to look at the developmental changes of speech processing. The findings are not conclusive. Some report that a small negative mismatch response (MMR), deemed to detect phonetic change (Nätänen, 2001), is observed in infants (Cheour et al., 2000; Friederici et al., 2002) and that this wave is similar to the one seen in adults (Dehaene-Lambertz and Gliga, 2004; Friederici, 2005). Other studies have found that the MMR switches from infancy to adulthood (Cheour et al., 2000; Cunningham et al., 2000; Sharma et al., 1997). For native language sounds, it appears that monolingual infants begin to make a switch from a positive response to a negative response somewhere around 11 and 13-months of age (Datta et al., 2010; Morr et al., 2002; Rivera-Gaxiola et al., 2005a, 2005b; Shafer et al., 2011)). Some evidence provided by fMRI studies also reports that 3-month old monolingual infants show bilateral activity in the superior temporal gyrus (STG) (Dehaene-Lambertz et al., 2002) just like monolingual adults (Binder et al., 2000) but other studies report that as monolingual infants transition into childhood, they undergo a process of interactive specialization whereby several distributed brain regions localize in the temporal lobe by 3-years of age (Redcay et al., 2008). Although Redcay’s findings suggest a later cutoff point of language-specific processing, this may simply be the result of the age groups studied. The inconsistencies found across studies may lie in that some analyses indirectly compare infants with adults while other studies look at the developmental changes that transpire year after year. Nonetheless, the developmental shift observed in neural processing appears to closely match the timeline described in behavioral studies thus reinforcing the notion that infants entrench neurally to their native language early on (Kuhl, 2000; Kuhl et al., 2008).

**Bilinguals**

Two recent EEG studies with simultaneous bilingual infants found that a negative response is elicited by contrasts from either language (Garcia-Sierra et al., 2011; Shafer et al., 2011). Shafer et al. (2011) presented the English contrast /ɛ/-/i/ (/ɛ/ as standard and /i/ as deviant) to 6–46 month old Spanish–English bilinguals and found that the 6-, 14-, and 29-month old infants had larger negative amplitude waves than their monolingual counterparts. The 30–46 month old bilinguals showed a response pattern that resembled that of monolinguals but with a later latency. The authors concluded that the amplitude of the negative response at different ages reflects the level of increased attention designated to speech perception. That is, younger bilinguals have larger responses than older bilinguals because they require more attentional resources to process speech. García-Sierra et al. (2011), however, did not find evidence of discriminability at 6 months of age in bilingual infants doing a double-oddball task (where the standard stimulus was a sound common to both languages and the deviant stimulus was a unique sound to each language), but did find that 12-month old bilingual infants could discriminate the phonemic change. This demonstrates that bilingual infants’ perception of both languages improves with age. These studies show that a positive mismatch response is not elicited in simultaneous bilingual infants the way it is elicited in monolingual infants. Therefore, monolinguals and simultaneous bilingual infants process native speech differently at the neural level.

**Speech learning models**

Typically, sequential learners are those who acquire the second language in adulthood. However, in the present study we have a unique sample of children who were exposed to the second language before the so-called “critical period” ended but after the first language had taken root. Therefore, we describe two behavioral models to explain how neural speech learning is expected to occur in these children. In the Perceptual Assimilation Model (PAM), Best (1994) and Best et al. (2001) proposes that new learners tend to assimilate the sounds of the second language (L2) into phonemic categories of the first language (L1), especially if the L2 phoneme highly resembles a L1 phoneme. In cases where the L2 phoneme is quite distinct from anything available in the L1 inventory, new learners create a new phonemic representation. It is worth noting that assimilating sounds is an advantageous strategy in the early stages of acquisition because it gives the learner the ability to communicate quickly. However, this temporary strategy can hurt long-term learning if the person does not eventually recruit the appropriate attentional resources to accurately discriminate novel sounds (Archila-Suerte et al., 2011; Guion et al., 2000). In the Speech Learning Model (SLM), Flege (1995, 2003) argues that the capacity to discriminate and categorize speech remains intact throughout life. Therefore, even late learners who acquire the second language in adulthood can learn to readjust the weights assigned to speech properties as long as there is allocation of attentional resources to phonetic information and a significant amount of L2 input. The exploratory nature of the present study compels us to connect speech-learning theory with evidence from speech-development studies, as the speech learning experiences of sequential bilingual children resemble that of adults but during a period of putatively greater brain plasticity.
Hypotheses

Since all children in our study learned Spanish as their native language at home and English as their second language at school, we expect children to be neurally committed to Spanish by the time they are exposed to English. Therefore, in the beginning stages of L2 acquisition – i.e., the learning stages bilinguals (6–8 yrs) will be undergoing – we expect the phonemic categories of the second language to be assimilated by the phonemic categories of the first language, as the perceptual assimilation model suggests. Neurofunctionally, this perceptual behavior should translate into activity of the same areas recruited to process L1 speech, i.e., the STG bilaterally. On the other hand, as children grow older and acquire more linguistic experience, we expect them to differentiate L2 speech more accurately via increased attention to phonetic information and increased explicit processes, as proposed by the speech-learning model. In other words, as bilingual children grow older and experienced in the phonology of the second language, higher-order cognitive mechanisms like attention, working memory, and cognitive control should be recruited to improve the discrimination of L2 sounds. Neurofunctionally, we expect to see activity in the parietal lobe for allocation of attention (Behrmann et al., 2004), and the prefrontal cortex for working memory and cognitive control (Hernandez et al., 2000, 2001; Owen et al., 2005).

Method

Participants

Two groups of subjects (N = 54) participated in this study: English monolingual children and Spanish–English bilingual children. The breakdown of subjects by group was the following:

English monolingual children (n = 13): All monolingual children were born in the United States and both parents spoke only English. The age range was from 6 to 10 years (M = 8.1, SD = 1.3). Only 1 child appeared to have been exposed to Spanish considerably; however, a language assessment demonstrated that the child’s knowledge of Spanish was negligible. All children attended local schools and had completed at least three years of formal education in English at the time of testing. Socioeconomic status (SES) was determined from the level of parental education using a 6-point likert scale (1 = Some or less than elementary education, 6 = Advanced degree). On average, the parents of monolingual children completed college-level education (M = 3.76, SD = 0.72).

Early Spanish–English bilingual children (n = 41): Three children who completed less than 60% of the fMRI task were dropped from analyses. The cutoff of 60% was used because these data did not have not enough vectors to be reliably analyzed. The rest of the children completed more than 80% of the task. All means and standard deviations reported here apply to the sample (n = 38). Bilingual children’s age range was from 6 to 10 years as well (M = 8.5, SD = 1.4). All children were exposed to Spanish from birth and the majority was exposed to English around 5 years of age (M = 4.5 years, SD = 2.2, range: 4 to 8, one child was exposed to English at 9). Overall language use per day across the contexts of school and home indicated that children used English 47% of the time and Spanish 52% of the time. About 90% of the children spoke Mexican Spanish. Only 3 children spoke Salvadorian Spanish and 1 child spoke Colombian Spanish. Of the 38 bilingual children, 28 were born in the United States and 10 were born in a Spanish-speaking country. All 10 immigrant children arrived in the United States before 5 years of age. The level of SES of bilingual children’s parents was significantly lower than monolingual children’s parents F (1,49) = 12.188, p < 0.001; that is, bilingual children’s parents received only some or less than a high school education (M = 2.05, SD = 1.06). For our analyses of interest regarding processing of non-native speech throughout childhood, we divided children into two groups: younger children (6 thru 8 years old; n = 16) and older children (9 thru 10 years old; n = 22).

English monolingual adults (n = 17): One subject was dropped from analyses due to technical issues with image acquisition. All subjects were born in the United States and had at least 15 years of education at the time of testing. The age range was from 24 to 35 years (M = 24.4, SD = 4.5). The parents of monolingual adults did not differ from the parents of monolingual children in socioeconomic status F (1,27) = 0.53, p > 0.05.

Early Spanish–English bilingual adults (n = 19): Like bilingual children, these participants were exposed to English around 5 years of age when formal schooling began. Spanish was the native language of all adult bilinguals. The average age of second language acquisition was 4.5 years (SD = 0.60). From childhood to adulthood, there was a significant increase in the use of English F (1,55) = 9.46, p = 0.003 and a significant decrease in the use of Spanish F (1,55) = 9.39, p = 0.003. Language use in adults was assessed across the contexts of school (i.e., college) and home; the same way this variable was assessed in children. Bilingual adults were comparable to monolingual adults in age (M = 21.8, SD = 3.6; F (1,54) = 3.7, p > 0.05) and years of education, with at least 15 years completed. However, the parents of bilingual adults had a significantly lower SES than the parents of monolingual adults F (1,33) = 10.613, p < 0.05 but did not have a significantly lower SES than the parents of bilingual children F (1,55) = 0.824, p > 0.05. Therefore, monolingual children and adults came from families of higher SES than bilingual children and adults.

Auditory stimuli

We presented the English syllables “saf,” “sof,” and “suf” (the corresponding phonetic notations are: /sæf/ as in hat, /s f/ as in hot, and /s f/ as in hut). We chose the fricatives /s/ and /f/ because these have been found to improve the recognition of the vowel (Fernandez et al., 1998) and the English vowels /æ/ / / / / because of their acoustic similarity with Spanish vowels. As demonstrated by Flege et al. (1995), experienced and inexperienced Spanish speakers of English rate these vowels as highly similar instances of the Spanish vowel /a/ (e.g., lata) (see Bradlow, 1995 for a detailed list of formant values of English and Spanish vowels). Selecting highly similar vowels across language makes the perceptual task more challenging, thus allowing us to better make the perceptual task more challenging, thus allowing us to better understand the neural mechanisms of non-native speech in children compared to adults.

Behavioral measures

Speech recording judgments

The speakers – children and adults – read 144 English words that targeted the production of the same vowel sounds heard during the fMRI task (/æ/, / /, / /). Seven young children who did not know
how to read were given a modified delayed imitation task. The performance between the children who did delayed imitation and those who read the words on their own did not differ on any of the vowels (/æ/ F (1,35) = 0.58, p = 0.81, / / F (1,35) = 1.6, p = 0.28, / / F (1,35) = 1.50, p = 0.22). Because the behavioral performance producing non-native speech between children who could read and children who could not read was similar, we did not expect the status of a higher-level skill (i.e., reading) to affect the results of a perceptual auditory task. All recordings were obtained in Praat with an external tabletop microphone (Omnidirectional Condenser, MX391/0). After the recordings, four English monolingual judges assessed the intelligibility of children’s and adults’ speech samples and the level of foreign accent of each speaker. The judges listened to all recordings presented in random order to prevent bias. To do this, judges transcribed what they heard and rated how noticeable foreign accent was on a 9-point scale (1 = No Foreign Accent, 9 = Very Strong Foreign Accent).

fMRI measures and procedure

Pre-attentive listening
In this block fMRI design, participants watched a muted movie of Planet Earth while the English syllables saf, sof, and suf played through a pair of scanner-compatible headphones. The movie played continuously during on-task and baseline trials. Participants were asked to focus their attention on the movie. No overt responses were collected. The syllables were presented in pairs during a silent interval between clustered image acquisitions. Each syllable pair belonged to a within-category or a between-category trial. Within-category trials contained syllables of the same phonemic category (e.g., saf-saf, sof-sof, and suf-suf) and between-category trials contained syllables of different phonemic categories (e.g., saf-sof, sof-suf, and suf-suf). Five consecutive trials composed a block of within- or between-category stimuli. A baseline block of silence was included to contrast the effects of speech. The total duration of the task was 27 min (see Fig. 1). In the analysis, all within- and between-category conditions were combined to form the overarching condition called “non-native speech” (or “native-speech”) for monolingual children) as the goal of the present study was to reveal how the neural mechanisms underlying overall non-native speech perception change throughout childhood. All participants were right handed and had no history of language disorders or speech-related difficulties.

Acquisition parameters

Whole-brain scans were performed with a 3.0 Tesla Magnetom Trio (Siemens, Germany) at the Human Neuroimaging Laboratory of Baylor College of Medicine in Houston, Texas. A total of 542 functional (T2*-weighted) images were acquired using a clustered volume acquisition (CVA) paradigm to quiet the scanner while the auditory stimuli were presented. An interleaved descending Echo Planar Imaging (EPI) sequence was employed with the following parameters: repetition time (TR) = 3 s, TR delay (silent interval) = 1.42 s, volume acquisition time (TA) = 1.58 s, transversal slices per volume = 26, TE = 30 ms; 5 mm thickness, 3.4 × 3.4 × 5.0 mm resolution, flip angle = 90°, with the centermost slice aligned with the anterior commissure and posterior cingulate (AC-PC). High-resolution anatomical images used a T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (TR = 1.2 s, TE = 2.66 ms, 1 mm3 isotropic voxel size) reconstructed into 192 slices. Auditory stimuli were presented using the in-house software NEMO (Network Experiment Management Objects). This software synchronized the task with the scanner with millisecond accuracy.

fMRI data analysis

Whole-brain analyses were conducted with SPM8 (Wellcome Trust Center for Neuroimaging, London) using a block design specification for our statistical model. Functional images were slice-time corrected, motion-corrected, aligned to anatomical scans, and normalized to MNI stereotactic space. Spatial smoothing used an 8 mm full-width half maximum Gaussian Kernel. Children’s data was further inspected for motion with the Artifact Detection toolbox (Whitfield-Gabrieli, 2010). Children’s brain scans that exceeded 1 mm of linear movement and 0.5° of angular movement were dropped from analyses. The range of scans dropped for each child subject varied from 0 to 140 scans. The

Table 1

<table>
<thead>
<tr>
<th></th>
<th>/saf/</th>
<th>/saf/</th>
<th>/saf/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dur</td>
<td>225</td>
<td>202</td>
<td>150</td>
</tr>
<tr>
<td>F1</td>
<td>761</td>
<td>721</td>
<td>650</td>
</tr>
<tr>
<td>F2</td>
<td>1414</td>
<td>1105</td>
<td>1227</td>
</tr>
<tr>
<td>F3</td>
<td>2525</td>
<td>2688</td>
<td>2726</td>
</tr>
</tbody>
</table>
average number of scans dropped was 84 and the total number of scans collected for each participant was 542; therefore, on average, about 15% of data points were dropped for each child. Relaxing the thresholds to 2 mm of linear movement and 2.8° of rotational movement did not change our results. Therefore, stringent movement thresholds did not lead to the loss of power in our data.

At the first level of analysis, the stimulus onsets of each condition (within-category, between-category, and baseline) were modeled with a canonical hemodynamic response function. For each subject, a new condition labeled “non-native speech” was created to investigate overall neural activity in response to non-native sounds by combining the effects of within- and between-category conditions. In our study, collapsing within-category and between-category conditions was an appropriate procedure at the first level of analysis because the two types of stimuli share a common attribute: they constitute speech. Modeling the new condition at the first level of analysis enabled us to compare between groups at the second level. The goal of combining conditions at the first level was to generate a pairwise contrast (non-native speech > baseline) that would allow us to compare the main conditions at the second level within each group and between groups of children and adults. The x,y,z coordinates reported in this study were automatically provided by SPM and the anatomical labeling was obtained from Anatomy Toolbox (Eickhoff et al., 2005). If there was ambiguity about the label of a particular region, we consulted Talairach Client v.2.4.2 (Lancaster et al., 2000) to reach a consensus.

Results

Behavioral

Children’s speech recording judgments

We analyzed 6 words that closely resembled the stimuli used in the fMRI pre-attentive listening task. The words were: gas, sad, lot, job, lunch, and cup. Consensus analyses indicated that when words were collapsed across phonemic categories (gas and sad for /æ/; lot and job for /ʌ/ and lunch and cup for /ɒ/), bilingual and monolingual children did not significantly differ in the production of /æ/ F (1,48) = 2.40, p = 0.12, /ʌ/ F (1,48) = 0.04, p = 0.95, or /ɒ/ F (1,48) = 1.26, p = 0.26. This indicates that both monolingual and bilingual children have distinct mental representations of the phonemes /æ/, /ʌ/, /ɒ/ and can discriminate between each category. Bilingual children’s foreign accent was significantly stronger than that of monolingual children but tended to improve with age and experience F (3,46) = 6.83, p < 0.001 (see Fig. 2).

Adult’s speech recording judgments

The same 6 words analyzed in children were analyzed in adults. Monolingual and early bilingual adults did not differ in the production of the sounds /æ/ F (1,30) = 1.25, p = 0.272, /ʌ/ F (1,30) = 0.77, p = 0.38, or /ɒ/ F (1,30) = 2.49, p = 0.12 and were rated as having similar non-accented pronunciation in English, F (1,30) = 1.17, p = 0.28 (see Fig. 2).

Neuroimaging results

Monolingual children

Left-lateralized activity of the STG was elicited by the contrast native speech > baseline (whole-brain FWE correction p < 0.05, peak z = 4.79, k = 15) in all monolingual children. Looking at each group of monolingual children, younger and older, we observed bilateral activity of the STG (R: z = 3.93, L: z = 4.46) in younger monolingual children (k = 15 voxels, threshold p < 0.001 whole-brain uncorrected, peak-level FWE p < 0.05 corrected) but not in older monolingual children. In comparing younger to older monolingual children, we found that listening to native speech sounds elicited more activity in early childhood than in late childhood in the parahippocampal gyrus (z = 4.71, peak-level FWE p < 0.05 corrected) and other right hemisphere areas (p < 0.001 uncorrected) including the fusiform gyrus (z = 4.14), thalamus (z = 2.95), and middle temporal gyrus (z = 2.82). No areas of activity were found in older monolingual children when compared to younger monolingual children (see Fig. 3).

Bilingual children

Bilateral activity of the STG (left z = 6.95, right z = 6.55) was elicited by the contrast non-native speech > baseline (whole-brain FWE correction p < 0.05, k = 15) in all bilingual children. Looking at each group of bilingual children, younger and older, we observed that listening to non-native speech elicited focused bilateral STG activity in younger bilingual children (R: z = 3.76, L: z = 3.92) but widespread activity in older bilingual children (p < 0.001 whole-brain uncorrected). Older children recruited the STG (R: z = 5.71, L: z = 6.17, p < 0.05 peak-level corrected), the superior parietal lobule – SPL (R: z = 4.01, L: z = 3.27), inferior parietal lobule – IPL (R: z = 3.78, L: z = 3.57), and the inferior frontal gyrus – IFG (R: z = 3.41, L: z = 3.46), bilaterally. Other clusters of activation were seen in the right superior medial gyrus – SMG (z = 3.38), and the right middle frontal gyrus – MFG (z = 4.12). (see Table 2 for a full description of the areas activated in children). A direct comparison between groups showed that older bilingual children engage the MFG (R: z = 2.85, L: z = 3.01) and IPL bilaterally (R: z = 2.75, L: z = 2.81), the left MTG (z = 2.83), and the right parahippocampal gyrus (z = 3.24) more than younger children (p < 0.005 uncorrected). The reversed contrast of younger bilingual children versus older bilingual children did not reveal any areas of activity (see Fig. 3). As seen here, brain activity survived FWE correction for the contrast of non-native speech > baseline when all 38 bilingual children were entered into the model but when smaller subgroups of younger and older bilinguals were investigated or when the groups of bilinguals were directly compared. Therefore, it is advisable to take the latter set of results with caution as there is a possibility for type 1 error. However, it is justifiable to use an uncorrected p-value for complex comparisons where the possibility of finding differences between groups is restricted by the similarity of the activation observed in each group independently.

Monolingual children versus monolingual adults

To further investigate how children process speech, we compared our group of interest (i.e., children) against adults (see Table 3 for a description of the areas found in children versus adults and vice versa). Here we found that young monolingual children not only show more activity than older monolingual children but also more activity than monolingual adults (p < 0.001 uncorrected). The areas of activity that survived in young monolingual children in comparison to monolingual adults were the left thalamus (z = 3.86), the right precentral gyrus (z = 3.58), and the right hippocampus (z = 3.53). On the other hand, monolingual adults showed more activity than monolingual older children but not more activity than young monolingual children. The areas observed in monolingual adults in comparison to older monolingual children were the hippocampus bilaterally (R: z = 4, L: z = 3.47), the right STG (z = 3.79), the right temporal pole (z = 3.67), and the left MTG (z = 3.65). This indicates that early childhood in monolinguals is a period of intense activity, late childhood is a period of low activity, and adulthood is a period of moderate activity (see Fig. 4).

Bilingual children versus bilingual adults

Young bilingual children did not show more activity than older bilingual children but did show slightly more activity in the right inferior occipital gyrus (z = 3.45) than bilingual adults. Older bilingual children, on the other hand, showed more activity than younger bilingual children and bilingual adults (p < 0.001 uncorrected). In comparison to the adults, older bilingual children showed activity in the right
Fig. 2. Speech recording judgments of children and adults. Panel A shows the production performance of children and adults across the phonemic categories /æ/, /ɛ/ and /ɜ/ according to judges’ consensus scores. Higher scores represent higher consensus. Panel B shows the degree of accent perceived in children and adults. Higher scores represent stronger foreign accent.

Fig. 3. Developmental trajectory of speech processing in monolingual and bilingual children. Monolingual children show more activity in early childhood in response to native speech. Bilingual children show more activity in late childhood in response to non-native speech.
IPL (z = 4.01), the left SPL (z = 4.22, peak-level FWE corrected < 0.05), bilateral cingulate gyrus (R: z = 3.53, L: z = 3.41), bilateral precen
tral gyrus (R: z = 3.65, L: z = 3.28), bilateral cerebellum (R: z = 3.39, L: z = 3.29), and left inferior temporal gyrus (z = 3.29). This suggests that the period of most intense and widespread activity in sequential bilinguals
occurs during late childhood, contrary to monolinguals that experience a period of more extensive activity during early childhood (see Fig. 4).

Speech processing across SES levels
A post-hoc question of interest was that of speech processing
given SES status in children. Since our sample of children belonged
to higher SES monolingual families or lower SES bilingual families, we conducted a 2 X 2 ANOVA analyses (monolingual, bilingual X high
SES, low SES) to investigate how children’s differences in SES influenced their
processing of speech. However, because monolingual children were
processing their native language and bilingual children were pro-
ceessing a non-native language, we did not compare the groups of mono-
linguals and bilinguals directly. We compared across SES but within
each group of children. In monolinguals, we found that high SES or
low SES does not influence how native speech is processed (p > 0.001
unc.); however, when a lower threshold of p < 0.005 unc. is applied, low SES monolingual children recruit the left MFC, an area known
for its involvement in cognitive control (Derrfuss et al., 2004). In biling-
uals, we found that high SES bilingual children recruit the right insula
(z = 3.41) when compared to low SES bilingual children (p < 0.001
uncorrected). In order to explore what other activity was potential in
the bilingual group, a lower threshold of p < 0.005 uncorrected was used
as well. Here, we found that in addition to activity in the right insula, high SES bilingual children activate the pallidum bilaterally, the left
rolandic operculum, left MTG, and right supramarginal gyrus. These
areas are reportedly involved in articulation (Kato et al., 2007), and detec-
tion of sound differences (Joanisse et al., 2007) (see Table 4 and Fig. 5).
Comparing neural activation across children and adults was not neces-
sary because adult’s SES equated children’s SES within each language group.

Dual processing: movie and speech
Having had children watch a silent movie and listen to speech syl-
lables at the same time raised the concern that the brain activity seen
in the contrasts reported above was the result of visual processing and
not auditory processing. To investigate this possibility, we fo-
cused on the overall activation pattern of the left STG and the right
parahippocampus — two areas known to be distinctively involved in
speech processing (Zevin and McCandliss, 2005; Zevin et al., 2010)
and visual processing of landscapes (Maguire et al., 2001). To do
this, we selected the coordinates of the left STG (−58 to −20) and
right parahippocampal gyrus (20 −46 to −12) from two previous stud-
ies that independently explored these phenomena i.e., Zatorre et al.
(1992) for speech and Maguire et al. (2001) for visual processing.
As seen in Fig. 6 panel A, monolingual children show increased
activity in the left STG during on-task trials of speech (subjects 1a
through 13a) and a sharp decrease in activity during baseline trials of
silence (subjects 1b to 13b). A reversed pattern is seen in the right
parahippocampal gyrus of monolingual children (panel B); there is
decreased activity during on-task speech trials and increased activity
during baseline trials. This indicates that monolingual children’s right
parahippocampal gyrus is slightly more sensitive to visual informa-
tion from the movie when there is no speech information available
(i.e., during baseline). But as seen in panels A and B, the switch
from speech to baseline conditions is less drastic in the right
parahippocampal gyrus than in the left STG. This may be due to the
fact that the movie continued playing during on-task trials

Table 2
Brain activity observed within and between groups of monolingual and bilingual children for the contrast speech > baseline.

<table>
<thead>
<tr>
<th>Hemisphere</th>
<th>Area</th>
<th>Voxels per cluster</th>
<th>Peak Z</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger monolingual children (6-8 yrs)</td>
<td>Superior temporal gyrus</td>
<td>745</td>
<td>4.46</td>
<td>−58 −10 2</td>
</tr>
<tr>
<td>Right</td>
<td>Superior temporal gyrus</td>
<td>897</td>
<td>3.39</td>
<td>60 −10 4</td>
</tr>
<tr>
<td>Older monolingual children (9-10 yrs)</td>
<td>Superior temporal gyrus</td>
<td>23</td>
<td>3.01</td>
<td>−46 −40 18</td>
</tr>
<tr>
<td>Younger bilingual children (6-8 yrs)</td>
<td>Superior temporal gyrus</td>
<td>247</td>
<td>3.92</td>
<td>−50 −22 10</td>
</tr>
<tr>
<td>Right</td>
<td>Superior temporal gyrus</td>
<td>417</td>
<td>3.76</td>
<td>62 −6 0</td>
</tr>
<tr>
<td>Older monolingual children (9-10 yrs)</td>
<td>Superior temporal gyrus</td>
<td>2082</td>
<td>6.17</td>
<td>−58 −22 8</td>
</tr>
<tr>
<td>Right</td>
<td>Superior temporal gyrus</td>
<td>2697</td>
<td>5.71</td>
<td>60 −12 6</td>
</tr>
<tr>
<td>Right</td>
<td>Middle frontal gyrus</td>
<td>272</td>
<td>4.12</td>
<td>54 −4 46</td>
</tr>
<tr>
<td>Right</td>
<td>Superior parietal lobule</td>
<td>306</td>
<td>4.01</td>
<td>34 −60 54</td>
</tr>
<tr>
<td>Right</td>
<td>Inferior parietal lobule</td>
<td>73</td>
<td>3.78</td>
<td>56 −38 50</td>
</tr>
<tr>
<td>Left</td>
<td>Inferior parietal lobule</td>
<td>269</td>
<td>3.57</td>
<td>−42 −44 42</td>
</tr>
<tr>
<td>Left</td>
<td>Inferior frontal gyrus</td>
<td>57</td>
<td>3.46</td>
<td>−36 10 30</td>
</tr>
<tr>
<td>Right</td>
<td>Inferior frontal gyrus</td>
<td>43</td>
<td>3.41</td>
<td>44 20 30</td>
</tr>
<tr>
<td>Right</td>
<td>Paracentral lobule</td>
<td>27</td>
<td>3.38</td>
<td>−6 −30 70</td>
</tr>
<tr>
<td>Left</td>
<td>Superior parietal lobule</td>
<td>34</td>
<td>3.27</td>
<td>−24 −64 52</td>
</tr>
<tr>
<td>Between children groups, speech &gt; baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemisphere</td>
<td>Area</td>
<td>Voxels per cluster</td>
<td>Peak Z</td>
<td>MNI coordinates</td>
</tr>
<tr>
<td>Younger monolingual children &gt; older monolingual children</td>
<td>Parahippocampal gyrus</td>
<td>256</td>
<td>4.71</td>
<td>34 −22 −18</td>
</tr>
<tr>
<td>Right</td>
<td>Fusiform gyrus</td>
<td>−</td>
<td>4.14</td>
<td>24 −20 −18</td>
</tr>
<tr>
<td>Right</td>
<td>Thalamus</td>
<td>82</td>
<td>2.95</td>
<td>4 −14 6</td>
</tr>
<tr>
<td>Right</td>
<td>Middle temporal gyrus</td>
<td>18</td>
<td>2.82</td>
<td>52 −2 −26</td>
</tr>
<tr>
<td>Older bilingual children &gt; younger bilingual children</td>
<td>Parahippocampal gyrus</td>
<td>18</td>
<td>3.24</td>
<td>18 −4 −18</td>
</tr>
<tr>
<td>Right</td>
<td>Middle frontal gyrus</td>
<td>22</td>
<td>3.01</td>
<td>−38 16 48</td>
</tr>
<tr>
<td>Left</td>
<td>Middle frontal gyrus</td>
<td>19</td>
<td>2.85</td>
<td>32 30 50</td>
</tr>
<tr>
<td>Left</td>
<td>Middle frontal gyrus</td>
<td>22</td>
<td>2.83</td>
<td>−52 −34 −6</td>
</tr>
</tbody>
</table>

Table 3
Brain activity observed between groups of children and adults within language groups of monolingual and bilingual.

| Hemisphere | Area                        | Voxels per cluster | Peak Z | MNI coordinates |
| Younger monolingual children > monolingual adults | Thalamus            | 51 | 3.86 | −6 −14 6 |
| Right      | Precentral gyrus            | 68 | 3.58 | 50 0 18 |
| Right      | Hippocampus                 | 17 | 3.53 | 32 −14 −20 |
| Monolingual adults > older monolingual children | Hippocampus         | 35 | 4.00 | 40 −34 −10 |
| Right      | Superior temporal gyrus     | 163 | 3.79 | 50 −14 10 |
| Right      | Temporal pole               | 42 | 3.67 | 52 6 −8 |
| Left       | Middle temporal gyrus       | 29 | 3.65 | −44 −4 −28 |
| Left       | Hippocampus                 | 25 | 3.47 | −18 −10 −8 |
| Younger bilingual children > bilingual adults | Inferior occipital gyrus | 32 | 3.45 | 36 −80 −4 |
| Older bilingual children > bilingual adults | Superior parietal lobule | 2806 | 4.22 | −18 −62 52 |
| Right      | Inferior parietal lobule    | −  | 4.01 | 28 −54 50 |
| Right      | Precuneus                   | 53 | 3.65 | 34 0 50 |
| Right      | Cingulate gyrus             | 70 | 3.53 | 20 2 32 |
| Right      | Cerebellar vermis           | 22 | 3.39 | 2 −44 −4 |
| Left       | Cerebellum                  | 24 | 3.29 | −70 −10 |
| Left       | Inferior temporal gyrus     | 21 | 3.29 | −46 −66 −6 |
| Left       | Precentral gyrus            | 22 | 3.28 | −34 −6 56 |

* significance of p < 0.005 whole-brain uncorrected.
when speech stimuli were presented. Therefore, it appears that visual processing of the movie remained constant throughout the task, although at lower intensity than speech processing. Like monolingual children, bilingual children in panel C also show increased activity in the left STG during on-task speech trials (subjects 1a to 38a) and decreased activity during baseline trials (subjects 1b to 38b). However, bilingual children’s pattern of activity in the right parahippocampal gyrus (panel D) is the same as in the left STG; that is, there is increased activity in the parahippocampal gyrus during on-task speech trials. This means that bilingual children engaged the parahippocampal gyrus to process auditory information instead of visual information.

Discussion
Our results demonstrate that the neural mechanisms supporting speech perception differ in monolingual and sequential bilingual children. The differences in neural processing between children of the same ages most likely result from the different types of linguistic experiences they undergo. The pattern of activation evoked by non-native speech sounds changes from early childhood to late childhood in bilingual children. While younger bilingual children engage perceptual areas of the temporal lobe to process speech, older bilingual children engage a

Table 4
Brain activity evoked by speech between groups of lower and higher socioeconomic status in monolingual and bilingual children.

<table>
<thead>
<tr>
<th>Socioeconomic status</th>
<th>Hemisphere</th>
<th>Area</th>
<th>Voxels per cluster</th>
<th>Peak Z</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual high SES &gt; bilingual low SES</td>
<td>Right</td>
<td>Insula</td>
<td>25 (445)</td>
<td>3.41</td>
<td>42 10 − 14</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>Rolandic operculum</td>
<td>121</td>
<td>3.02</td>
<td>− 42 6 2</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>Pallidum</td>
<td>171</td>
<td>2.93</td>
<td>− 10 − 8 − 6</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>Supramarginal gyrus</td>
<td>55</td>
<td>2.88</td>
<td>60 − 20 24</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>Pallidum</td>
<td>116</td>
<td>2.86</td>
<td>10 − 8 − 4</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>Middle temporal gyrus</td>
<td>58</td>
<td>2.8</td>
<td>− 64 − 44 4</td>
</tr>
</tbody>
</table>

| Monolingual high SES > monolingual low SES | None | Middle frontal gyrus | 36 | 2.87 | − 32 52 16 |

Note: Italicized area p-value < 0.001 uncorrected.

Fig. 4. Comparisons of speech processing between children and adults. In monolinguals, younger children show more activity than adults but adults show more activity than older children. In bilinguals, both younger and older children show more activity than adults.

A) High SES > Low SES in bilingual children, p < 0.001 unc.

B) Low SES > High SES in monolingual children, p < 0.005 unc.

C) High SES > Low SES in bilingual children, p < 0.005 unc.

Fig. 5. A. Bilingual high SES > bilingual low SES. Bilingual children from better socioeconomic status recruit the right Insula, p < 0.001. B and C) Exploratory analyses with an uncorrected p value of 0.005. Panel B shows the results of the contrast monolingual low SES > monolingual high SES. Panel C shows the results of the contrast bilingual high SES > bilingual low SES. The reverse contrasts did not elicit any activity in high SES monolingual or low SES bilingual children.
network of areas in the frontal and parietal lobe. These findings support two behavioral models of speech learning; namely, the PAM (Best, 1994; Best et al., 2001) and the SLM (Flege, 1995, 2003). That is, younger inexperienced bilingual children assimilate sounds of the second language onto first language phonemic categories via recruitment of the same brain areas observed in monolingual children, i.e., the STG, because assimilation is an advantageous strategy to begin learning a new sound system (Flege, 2003; Flege and Mackay, 2004; Flege et al., 1999). As bilingual children gather more experience with the phonology of L2, increased attention and working memory resources – via recruitment of inferior parietal lobe and middle frontal gyrus – are allocated to phonetic information, thereby improving older bilingual’s perception and production of L2.

Our neuroimaging results show that monolingual children recruit the STG in early and late childhood to process native speech. The STG coordinates found in monolingual children closely match the STG coordinates reported in other speech perception studies with monolingual adults (Binder et al., 2000). But it appears that monolingual children initially process speech information bilaterally and gradually lateralize this activity to the left hemisphere. Comparing groups of monolingual children we found that younger children showed activity in more areas than older children, consistent with Redcay et al.’s (2008) conclusion of interactive specialization for the processing of speech in monolingual children. Here we found that younger monolingual children showed more activity in several right-hemisphere areas including the thalamus, the middle temporal gyrus, the fusiform gyrus, and the parahippocampal gyrus. This mixture of speech-related and visual-processing areas may be the result of increased connectivity in the brains of younger monolingual children (Beaulieu et al., 2005; Suppékar et al., 2009).

Sequential bilingual children showed a different pattern, however. While younger bilingual children recruited the same brain regions known to be involved in the processing of native sounds in monolinguals (Binder et al., 2000), namely, the STG bilaterally, older bilingual children recruited a distributed network of brain areas (i.e., bilateral superior temporal gyrus, right middle frontal gyrus, bilateral superior and inferior parietal lobule, and bilateral inferior frontal gyrus). The coordinates of these areas are proximal to those found in other studies and have been linked to the early analysis of speech (Binder et al., 2000; Joanisse et al., 2007), working memory (Carlson et al., 1998; Casey et al., 1998; Martinkauppi et al., 2000; Owen et al., 2005), allocation of attention (Bushara et al., 1999; Liu et al., 2003; Zatorre et al., 1999), and articulatory processes (Bookheimer, 2002; Gelfand and Bookheimer, 2003; Hillis et al., 2004). A comparison across age groups in bilingual children showed that older bilingual children have increased activity in the inferior parietal lobule and middle frontal gyrus bilaterally, right parahippocampal gyrus, and left middle temporal gyrus. It is possible, therefore, that younger bilingual children exclusively recruit the STG because activity in this area facilitates learning of L2 phonemes in the early stages of phonological acquisition via assimilation of L2 speech sounds onto L1 phonology, as the PAM would suggest. However, as bilingual children grow older and experienced in the phonology of the second language, higher-order cognitive areas are recruited to improve the discrimination of L2 sounds. As noted by the SLM, increasing attentional resources to phonetic information helps to enhance novel L2 speech learning in...
sequential bilingual adults. Our results reveal that such recruiting of attentional and higher-order processes starts around 9 years of age in sequential bilinguals as these children significantly rely on attention (inferior parietal lobule), and working memory (middle frontal gyrus) more than younger children to discern phonetic information (middle temporal gyrus) from non-native sounds. Studies have found that successful maintenance of working memory involves a frontoparietal network in which increased attention to task-relevant information helps to update the contents of working memory (Roth and Courtney, 2007; Roth et al., 2006). Based on this information we conclude that even though the task did not require subjects to overtly attend to the sounds, older bilingual children increased their attention to speech to manipulate and retain non-native phonological information.

In our behavioral results, we learned that both monolingual and bilingual children have phonemic representations of the English sounds /a/, /o/, and /u/ because they can produce words that target the articulation of these vowels in a fairly accurate manner. However, bilingual children start out with L2 phonemic representations that readily map onto first language phonology, thus resulting in foreign accents. In time, children's variability in the perception of sounds reaches stability in adulthood when monolinguals and bilinguals become indistinguishable in their performance. Taking our behavioral and neuroimaging results together, we note that the brains of monolingual and sequential bilingual children process speech information differently at the neural level despite their similarities in behavior. Both monolinguals and bilinguals improve their perception of speech over time but they each accomplish the task in different ways. Monolinguals' main area for speech processing is the STG but there is also specialization of activity from early childhood to late childhood (Redcay et al., 2008). On the other hand, sequential bilinguals recruit the STG but mainly as a parasitic area that facilitates the initial processing of non-native speech. In line with our findings, studies of phonetic training have found the STG to be the locus of plasticity from which new learning departs (Callan et al., 2003; Golestani and Zatorre, 2004; Tricomi et al., 2006). In addition, instead of specializing to particular areas in late childhood, sequential bilinguals call on executive areas to continue improving the perception of non-native speech.

Monolingual younger children not only show more activity than older children but also more activity than adults, thus indicating that the brains of young monolinguals spend a significant amount of neural resources on the perception of speech. More importantly, monolingual adults showed more activity in particular regions of the STG, MTG, and hippocampus than older children. Therefore, late childhood in monolinguals is a relatively quiet period for the processing of speech. Perhaps older children's brains are better equipped to process other types of information besides speech. The activity seen in the STG and anterior MTG of monolingual adults may reflect participants' readiness to map meaning onto sound. That is, hearing the syllables “saf, sof, and suf” possibly primed the nodes of similar-sounding words. This conclusion is supported by recent studies that attribute the function of sound-semantic mappings to the anterior pole of the left middle temporal gyrus (Crinion et al., 2003; Tsapkini et al., 2011).

In bilinguals, both younger and older children showed more activity than adults. It is especially interesting to see that late childhood in sequential bilingual children continues to be the time period with greatest activity. Unlike older monolingual children who are undergoing a period of intermission for speech, older bilingual children are allocating much neural capacity for the processing of non-native speech. Besides the attention areas of the inferior and superior parietal lobe (Hugdahl et al., 2000), older bilingual children show more bilateral activity in the precentral gyrus, cingulate gyrus, and cerebellum than bilingual adults. These areas have been respectively associated with articulation (Baldo et al., 2011; Pulvermuller et al., 2006), error monitoring (Garavan et al., 2003; Kiehl et al., 2000), and timing of articulatory sequences (Ackermann and Hertrich, 2000). Therefore, late childhood is the time when older sequential bilingual children are using compensatory control mechanisms in order to distinguish sounds in the second language. Given that adult bilinguals do not show greater activity than older children, we know that the period of increased brain activity in late childhood recedes or specializes sometime before early adulthood. In a future study, we hope to replicate the results obtained here using an active listening task.

It is possible that some of the differences observed between monolinguals and bilinguals are attributed to differences in SES, as SES has been found to correlate with the level of phonological awareness (Stewart, 2004). The results of the 2 × 2 ANOVA, where the high SES group consisted of children whose parents had at least completed more than a high school education and the low SES group consisted of children whose parents had less than a high school education, showed that a high or low socioeconomic status does not moderate brain activity in monolingual children. Since monolinguals are exposed to speech since infancy, it is possible that overexposure to the sounds of their native language overrides any SES-driven disadvantages. However, further exploratory results with a lower threshold showed activity showed activity in the left MFG in low SES monolingual children, an area of cognitive control (Derrfuss et al., 2004), possibly indicating that low SES in monolinguals can cause a disadvantage if lack of educational resources are severe and persistent. In bilingual children, a high SES level elicits increased activity in the right insula, bilateral pallidum, left Rolandic operculum, left MTG, and right supramarginal gyrus. Given the insula’s and Rolandic operculum’s involvement in speech articulation (Baldo et al., 2011; Kato et al., 2007) and the right supramarginal gyrus’s involvement in fine-grained speech perception (Joanisse et al., 2007) it appears that higher SES bilingual children exploit their speech perception and articulatory processes better than low SES bilingual children. These post-hoc results, though intriguing and worth investigating more in-depth in the near future, must be taken with some caution as the sample of monolingual and bilingual children in each respective SES group was very small (monolingual low SES n = 5, monolingual high SES n = 8, bilingual low SES n = 28, bilingual high SES n = 10).

Finally, in understanding how children pre-attentively perceived visual and auditory information, we looked at the pattern of activation of the left STG and the right parahippocampal gyrus throughout the task. Here we found that monolingual children show more activity in the left STG when speech sounds were playing and more activity in the right parahippocampal gyrus when the speech sounds were not playing and only visual stimulation was present. This most likely reflects a switch between visual and auditory processing in monolingual children. But as seen in the results, the switch from speech to baseline is not as abrupt in the right parahippocampal gyrus because the movie continued to play when speech stimuli were presented. Bilingual children had more activity in the left STG and right parahippocampal gyrus when speech sounds were playing. While the STG is expected to be activated in response to speech, the parahippocampal gyrus is not – as this area is normally involved in visual processing. Therefore, it appears that bilingual children recruit an area intended for visual processing to facilitate their perception of non-native speech. The pattern of activation observed in the right parahippocampal gyrus was also observed in the right inferior occipital gyrus (IOG) of bilingual children. Based on this, it is likely that activity in the right parahippocampal gyrus evoked by the contrast older bilingual children > younger bilingual children and the activity in the right IOG evoked by the contrast younger bilingual children > bilingual adults is in response to speech, and not visual stimulation. A potential limitation to our findings is that children were not asked for details about the movie after their fMRI session. Hypothetically, however, if monolingual and bilingual children had been asked to retell what they saw in the movie, bilingual children should have been less able to provide details as one of the brain regions designated to process visual information was being recruited to process speech information.
Consequently, bilingual children should have had fewer resources available to process the movie. This small detail of the methodology will be taken into consideration in future studies. Lastly, a brief post-hoc examination of activity in the left STG and right parahippocampal gyrus as a function of SES showed that both monolingual and bilingual children who come from high SES backgrounds activate these areas more intensely than children who come from low SES backgrounds. Although the question of SES is not the main goal of the study, these preliminary results seem to suggest that children from higher SES backgrounds are stronger activators than children from low SES backgrounds. An interesting future question is how intensity of brain activity correlates with children’s proficiency of the first and second language.

This is the first study that investigates the neural basis of non-native speech perception in sequential bilingual children. By studying a group of 6- to 10-year-old children who were all exposed to the non-native language around 5 years of age, we were able to examine the developmental changes associated with the processing of speech in children transitioning from inexperienced to proficient users of L2 phonology. Given that the groups of monolingual and bilingual children were matched by age, we attribute the differences in neural activity to the unique phonological experiences sequential bilingual children go through. The most significant changes in the developmental trajectory of speech learning in monolingual children occur in early childhood, as demonstrated here, and in early infancy (Datta et al., 2010; Morr et al., 2002). On the other hand, the most significant changes in the developmental trajectory of speech learning in sequential bilingual children occur in late childhood. Sequential bilingual children develop alternative ways for processing non-native speech. In early childhood, perception of non-native sounds depends on the recruitment of perceptual areas to assimilate L2 sounds onto L1 categories, as predicted by the PAM; but as children gain experience with the phonology of the second language, higher-order brain areas are engaged to improve how non-native speech is perceived and articulated, as predicted by the SLM.

Our findings support the two behavioral models of speech learning introduced in the literature review, the PAM and the SLM. Experience with the phonology of the second language changes how the brains of sequential bilingual children process speech information throughout childhood. A shift in processing departs from sensory/perceptual areas already engaged in native speech perception to higher-order executive areas that help refine the perceptual details of nonnative speech. The analogous behavioral process, as proposed by the PAM and SLM, is employing an assimilation strategy in the early stages of L2 acquisition and a heightened-attention strategy in later stages to improve perception of phonetic details. Therefore, the results in the present study show that the neural system processes phonological information differently depending on the stage of L2 speech learning. While younger inexperienced bilingual children assimilate L2 speech via the bilateral STG, older more experienced bilingual children increase their attention to sounds and make use of their working memory to retain phonological information via the bilateral inferior parietal lobe and middle frontal gyrus. Activating a network of executive functions helps older bilingual children enhance their perception of non-native sounds. Although executive function begins to develop in early childhood (Marchovitch and Zelazo, 2009; Zelazo et al., 2003), recent evidence from Bialystok et al. (2006, 2011), Bialystok et al. (2008, 2009) suggests that the experiences generated by bilingualism have a positive effect on linguistic and cognitive processing. Our results support the notion that experience with non-native speech improves other aspects of cognitive processing as the recruitment of executive brain regions in older bilingual children is an alternative way to manipulate perceptual information (Archila-Suerte et al., 2011; Okamoto et al., 2007; Tallal and Gaab, 2006). An area of future study is the continued development of non-native speech perception in adolescent bilinguals who have been exposed to the second language sequentially. Investigating this will elucidate the ongoing neural processes of specialization or distributed processing in bilingual populations.

Conclusion

In summary, both monolingual and sequential bilingual children demonstrate behavioral improvement in their perception of English but each group does so recruiting different neural mechanisms. Our results suggest that experience with the sounds of the second language drives the effects seen in our sample of bilingual children. While lack of experience with L2 phonology leads to the recruitment of sensory/perceptual areas associated with the processing of native speech, experience with non-native sounds leads to the recruitment of executive brain regions to manipulate perceptual information.

Acknowledgments

We thank Tom Zeffiro from the Martins Center for Biomedical Imaging at the Massachusetts General Hospital for his guidance on fMRI analysis and Ferenc Bunta for his advice on the literature search. We also want to thank the team of research assistants who helped collect the data for this project: Magdalena Benavides, Heather Dial, Thalia Flores, Ashley Mendez, Gabriela Ochoa, Daniela Perez-Silva, and Jessica Romero. This work was supported by the Institute for Biomedical Imaging Science (IBIS) for Plasticity in Speech Perception in Early Bilingual Children and grant number R21HD059103-01 for the Neural correlates of lexical processing in child L2 learners from the National Institute of Health (NIH).

References


