Developmental Changes in Brain Activation Involved in the Production of Novel Speech Sounds in Children

Hiroshi Hashizume,[1]* Yasuyuki Taki,[1,2,3] Yuko Sassa,[1] Benjamin Thyreau,[2] Michiko Asano,[1] Kohei Asano,[1] Hikaru Takeuchi,[1] Rui Nouchi,[4,5] Yuka Kotozaki,[5] Hyeonjeong Jeong,[6] Motoaki Sugiura,[4,6] and Ryuta Kawashima[1,5,6]

[1]*Division of Developmental Cognitive Neuroscience, Institute of Development, Aging and Cancer, Tohoku University, Sendai, 980-8575, Japan*
[2]*Division of Medical Image Analysis, Department of Community Medical Supports, Tohoku Medical Megabank Organization, Tohoku University, Sendai, 980-8575, Japan*
[3]*Department of Radiology and Nuclear Medicine, Institute of Development, Aging and Cancer, Tohoku University, Sendai, 980-8575, Japan*
[4]*International Research Institute of Disaster Science, Tohoku University, Sendai, 980-8579, Japan*
[5]*Smart Ageing International Research Centre, Institute of Development, Aging and Cancer, Tohoku University, Sendai, 980-8575, Japan*
[6]*Department of Functional Brain Imaging, IDAC, Tohoku University, Sendai, 980-8575, Japan*

◆━━━━━━━━━━━━━━━━━━━━━━━━━━━━◆

Abstract: Older children are more successful at producing unfamiliar, non-native speech sounds than younger children during the initial stages of learning. To reveal the neuronal underpinning of the age-related increase in the accuracy of non-native speech production, we examined the developmental changes in activation involved in the production of novel speech sounds using functional magnetic resonance imaging. Healthy right-handed children (aged 6–18 years) were scanned while performing an overt repetition task and a perceptual task involving aurally presented non-native and native syllables. Productions of non-native speech sounds were recorded and evaluated by native speakers. The mouth regions in the bilateral primary sensorimotor areas were activated more significantly during the repetition task relative to the perceptual task. The hemodynamic response in the left inferior frontal gyrus pars opercularis (IFG pOp) specific to non-native speech sound production (defined by prior hypothesis) increased with age. Additionally, the accuracy of non-native speech sound production increased with age. These results provide the first evidence of developmental changes in the neural processes underlying the production of novel speech sounds. Our data further suggest that the recruitment of the left IFG pOp during the production of novel speech sounds was possibly enhanced due to the mat-

uration of the neuronal circuits needed for speech motor planning. This, in turn, would lead to improvement in the ability to immediately imitate non-native speech. *Hum Brain Mapp 35:4079–4089, 2014.* © 2014 Wiley Periodicals, Inc.

Key words: development; cognitive neuroscience; childhood; speech production; second language; fMRI

◆　══　◆

INTRODUCTION

Learning to speak a new language usually requires acquisition of skills to verbally produce non-native (L2) speech sounds. Childhood is known to be the best time in life to acquire L2 speech sounds. However, several behavioral studies have shown that adults can produce L2 speech sounds more accurately than children can in the early stage of learning L2 [Aoyama et al., 2004; Oh et al., 2011]. Furthermore, the other studies have reported that older children are more successful at producing L2 speech sounds than younger children are during the initial stages of learning, until after about 4–5 months of immersion exposure [García and García, 2003; Olson and Samuels, 1973; Snow and Hoefnagel-Höhle, 1977]. Given the prevailing idea of the advantage for children, especially younger children in learning pronunciation of L2 speech sounds [Piske et al., 2001], it is important to understand how and why the ability to immediately produce L2 speech sounds evolves as age increases. In this study, we attempted to reveal the neuronal underpinning of the age-related increase in the accuracy of L2 speech production in the early stage of learning. We expected that it may be attributable to the development of the brain regions involved in speech production control, such as speech motor planning, and to the differential recruitment of those brain regions according to age.

To elucidate the cognitive development of children, a number of developmental neuroimaging studies have examined age-related changes in brain activation during various tasks including language and cognitive control [Bitan et al., 2007; Booth et al., 2003; Brown et al., 2005; Bunge et al., 2002; Schapiro et al., 2004; Tamm et al., 2002]. The developmental increase in activation during various language tasks such as rhyming judgment and single word generation have been reported in the left inferior frontal gyrus (IFG) as well as the left temporo-parietal regions [Bitan et al., 2007; Brown et al., 2005; Schapiro et al., 2004]. Those age-related increase of the task-specific activation, especially in frontal regions, have been associated with a developmental specialization of task-relevant brain function [Bitan et al., 2007; Brown et al., 2005; Schapiro et al., 2004; Tamm et al., 2002]. However, age-related changes in brain activity during the production of L2 speech sounds have not yet been clarified.

In adults, a number of functional imaging studies have examined speech production using L1 speech sounds and found that prefrontal regions, such as the posterior part of the left IFG (Broca's area), had a key role in the speech motor planning for speech production [Bohland and Guenther, 2006; Guenther et al., 2006; Papoutsi et al., 2009; Riecker et al., 2008; Soros et al., 2006]. In terms of bilingualism, although L1 and L2 speech production are hypothesized to recruit the same neural networks, such as sensorimotor control systems, L2 speech production is less automatic and efficient than is L1 speech production, which results in greater activation during L2 than L1 [Simmonds et al., 2011a]. Moser et al. [2009] investigated brain activation during the production of unfamiliar, sublexical L2 speech sounds in normal adults, focusing specifically on the left IFG pars opercularis (pOp) and anterior insula (aIns), two primary areas of damage in apraxia of speech [Broca, 1865; Dronkers, 1996; Hillis et al., 2004; Mohr et al., 1978; Nagao et al., 1999; Ogar et al., 2006]. The authors found greater activation in those regions during the production of L2 compared with L1 speech sounds. Therefore, they hypothesized that the left IFG pOp and aIns are recruited to a greater extent in L2 speech sounds as unfamiliar L2 speech sounds do not exist in one's repertoire and require new motor planning. Recently, the greater activation in the left posterior part of the IFG as well as in the other sensorimotor regions during L2 speech production by adult speakers who became bilingual later in life has also been confirmed [Simmonds et al., 2011b].

Therefore, the purpose of this study was to identify age-related changes in brain activation during L2 speech production by children as well as to examine age-related changes in the accuracy of L2 speech production. We used functional magnetic resonance imaging (fMRI) as well as voice recordings during fMRI scanning of children aged 6–18 years. We were primarily interested in the left IFG pOp and aIns, which have shown greater activation during L2 than L1 speech production in adults [Moser et al., 2009]. We hypothesized that activation in the left IFG pOp and/or aIns would increase during the production of novel relative to that of native (L1) speech sounds and that the accuracy of L2 speech production would increase with age.

MATERIALS AND METHODS

Participants

Fifty-two healthy, right-handed, native Japanese children aged 6–18 years participated in these fMRI experiments as

part of our research project on the brain development of healthy children [Taki et al., 2011a, 2011b]. Fourteen participants were excluded from the analyses: 10 due to excessive head motion (greater than 3° of rotation or greater than 3 mm of translation) and four due to poor task performance during the fMRI session (more details are provided in the Results section). Even after the exclusion above, there were 18 subjects that still had large movement that was >1° or 1 mm. Then, to eliminate the effect of the volume with high motion, we adopted a similar method to the "scrubbing" approach, calculating framewise displacement [Power et al., 2012], which is mentioned below. However, since 30% of the data from one subject was excluded, compared to <10% exclusion rate of other subjects, all data from this subject was excluded for the study. Thus, the present study included data from 37 participants (18 males and 19 females). The age distribution is shown in Table I. All had normal vision, and none had a history of neurological or psychiatric illness. Handedness was evaluated using the Edinburgh Handedness Inventory [Oldfield, 1971]. All subjects were monolingual Japanese speakers. Under the Japanese educational system, participants older than 9 had received English instruction; however, none had any other significant second-language exposure (e.g., staying abroad longer than 1 month or attending foreign-language schools). Written informed consent was obtained from each participant and a parent. The study met all criteria for approval by the Institutional Review Board of the Tohoku University Graduate School of Medicine.

Stimuli

We used a total of 32 syllables that consisted of a consonant and a vowel (CV) as L2 and L1 speech sound stimuli. We collected possible CV combinations for the L2 from several languages (Persian, Korean, Spanish, and French) that were expected to be unfamiliar to our participants. We collected possible L1 combinations from Japanese, and these were expected to be familiar to the participants. Next, we selected the 16 most "unfamiliar" CVs for L2 (Table II) and the 16 most "familiar" CVs for L1 (Table III) from the CV combinations by administering questionnaires to 10 Japanese adult participants (who differed from the participants in the fMRI experiment) to determine the extent to which the sounds were dissimilar to Japanese CVs. We inadvertently included 1 CCV sound in the 16 L2

CVs. Each syllable was spoken by one female and one male adult native speaker of the language and was recorded digitally in a soundproof room at a sampling rate of 44,100 Hz. All sound stimuli were normalized so that those peak levels were close to 0 dB. The average length and standard deviation of the speech sounds were approximately 382 ± 106 ms. The average lengths of L2 and L1 were 396 and 368 ms, respectively, which were not significantly different ($P = 0.30$). The speech sounds were presented through a pair of scanner-compatible headphones (Resonance Technology, Northridge, CA, USA) during MRI scanning.

Tasks and Procedures

To investigate brain activation during the production of unfamiliar, L2 speech sounds, we conducted a block-designed fMRI in which participants performed an overt repetition task. Participants had to listen to a L1 or L2 target speech sound first (the CV was repeated two times within a single trial by different speakers) and then overtly repeat it by attempting to imitate it. To control for the additional perceptual difficulty of L2 sounds, we also included a perceptual control task. In this perception task, participants heard two successive speech sounds from different speakers and had to judge and verbally report whether the two sounds represented the same syllable. Here, the use of a discrimination task and the requirement of a verbal response were intended to maintain the subject's attention and control the length of utterances, respectively. Overall, our fMRI experiment comprised a two-by-two factorial design of language and tasks: repetition in L1 (R1), repetition in L2 (R2), discrimination in L1 (D1), and discrimination in L2 (D2).

Each task block (20 sec) was followed by a 12-sec rest period, and consisted of four trials, each of which was 5 sec in duration. During a trial, two sequential auditory stimuli (each 0.7 sec) were presented, followed by the participants' responses (2.1 sec) (Fig. 1). The task of the upcoming trial (repetition or discrimination) was presented visually to the participants during each trial. The language of the auditory stimuli (L1 or L2) was not presented.

In the repetition block, the same two syllables spoken by different speakers (a male and a female) were presented sequentially in a trial (Fig. 1a). The order in which the male/female voices were presented was pseudorandomized in each block. In each trial, immediately after listening to the pair of speech sounds, participants were asked to overtly repeat the syllable that they had heard, reproducing the sound's phonological features as closely as possible (imitation).

In the discrimination block, two of the same or different syllables spoken by different speakers were presented sequentially in a trial (Fig. 1b). The probability of same/different was 50%. In each trial, participants were asked to judge whether the pair was phonologically the same by

TABLE I. Age distribution of subjects included in analyses

Age (years)	7–9	9–11	11–13	13–15	15–17	17–19
Male	3	4	5	4	2	0
Female	3	3	6	6	0	1
Total	6	7	11	10	2	1

TABLE II. Non-native syllables (L2) used in the study

IPA	Language	Manner	Place	Voice	Vowel
ɲu	Spanish	Nasal	Palatal	Voiced	Oral close back rounded
xa	Spanish	Fricative	Velar	Voiceless	Oral open central unrounded
ra	Spanish	Trill	Alveolar	Voiced	Oral open central unrounded
ri	Persian	Trill	Alveolar	Voiced	Oral close front unrounded
fi	Persian	Fricative	Labiodental	Voiceless	Oral close front unrounded
ʒi	Persian	Fricative	Post alveolar	Voiced	Oral close front unrounded
dʒu	Persian	Affricate	Alveolar/lateral	Voiced	Oral close back rounded
ɣu	Persian	Fricative	Velar	Voiced	Oral close back rounded
ve	Persian	Fricative	Labiodental	Voiced	Oral close-mid front unrounded
xe	Persian	Fricative	Velar	Voiceless	Oral close-mid front unrounded
c'i	Korean	Fricative	Palatal	Voiceless	Oral close front unrounded
co	Korean	Plosive	Palatal	Voiceless	Oral close-mid back rounded
vu	French	Fricative	Labiodental	Voiced	Oral close back rounded
ʃu	French	Fricative	Post alveolar	Voiceless	Oral close back rounded
ʒu	French	Fricative	Post alveolar	Voiced	Oral close back rounded
ʁa	French	Approximant	Uvular	Voiced	Oral open front unrounded
IPA: International phonetic alphabet					

providing a predetermined overt response of "O" or "Chi," which are the first syllables of the Japanese words meaning "same" or "different," respectively. Overt responses were used under all conditions to balance task demands. The predetermined rule of "O/Chi" was easy for native Japanese speakers to remember.

The use of the two different speakers for each syllable was intended to highlight the novel phonetic features of the stimuli rather than such speaker-dependent factors as tone of voice. Under the R2 condition, participants were supposed to repeat speech sounds based on their common phonetic features; under the D2 condition, they were supposed to discriminate speech sounds by determining whether a common phonetic feature existed.

The voices of the participants during the experiment were recorded using an MR-compatible noise-cancelling microphone (FOMRI II, Optoacoustics, Israel). To minimize head-motion-related artifacts, participants' speech was produced during the silence between scans (see the next subsection for further explanation of the *sparse* scan), and pads were used to carefully fix participants' heads in place in the scanner. The production during the silence between scans also allowed clear recordings of the participants' utterances.

Image Acquisition

All MRI data were acquired with a 3-T Philips Intera Achieva scanner. Twenty-six gradient-echo echo planar

TABLE III. Native syllables (L1) used in the study

| IPA	Language	Manner	Place	Voice	Vowel
ka	Japanese	Plosive	Velar	Voiceless	Oral open central unrounded
ku	Japanese	Plosive	Velar	Voiceless	Oral close back/central rounded
ko	Japanese	Plosive	Velar	Voiceless	Oral close-mid back rounded
sa	Japanese	Fricative	Alveolar	Voiceless	Oral open central unrounded
so	Japanese	Fricative	Alveolar	Voiceless	Oral close-mid back rounded
tu	Japanese	Affricate	Alveolar	Voiceless	Oral close back/central rounded
te	Japanese	Plosive	Dental	Voiceless	Oral close-mid front unrounded
to	Japanese	Plosive	Dental	Voiceless	Oral close-mid back rounded
hi	Japanese	Fricative	Palatal	Voiceless	Oral close front unrounded
he	Japanese	Fricative	Glottal	Voiceless	Oral close-mid front unrounded
mo	Japanese	Nasal	Bilabial	Voiced	Oral close-mid back rounded
jo	Japanese	Approximant	Palatal	Voiced	Oral close-mid back rounded
ru	Japanese	Flap	Alveolar	Voiced	Oral close back/central rounded
gu	Japanese	Plosive	Velar	Voiced	Oral close back/central rounded
zu	Japanese	Affricate	Alveolar	Voiced	Oral close back/central rounded
da	Japanese	Plosive	Dental	Voiced	Oral open central unrounded
IPA: International phonetic alphabet					
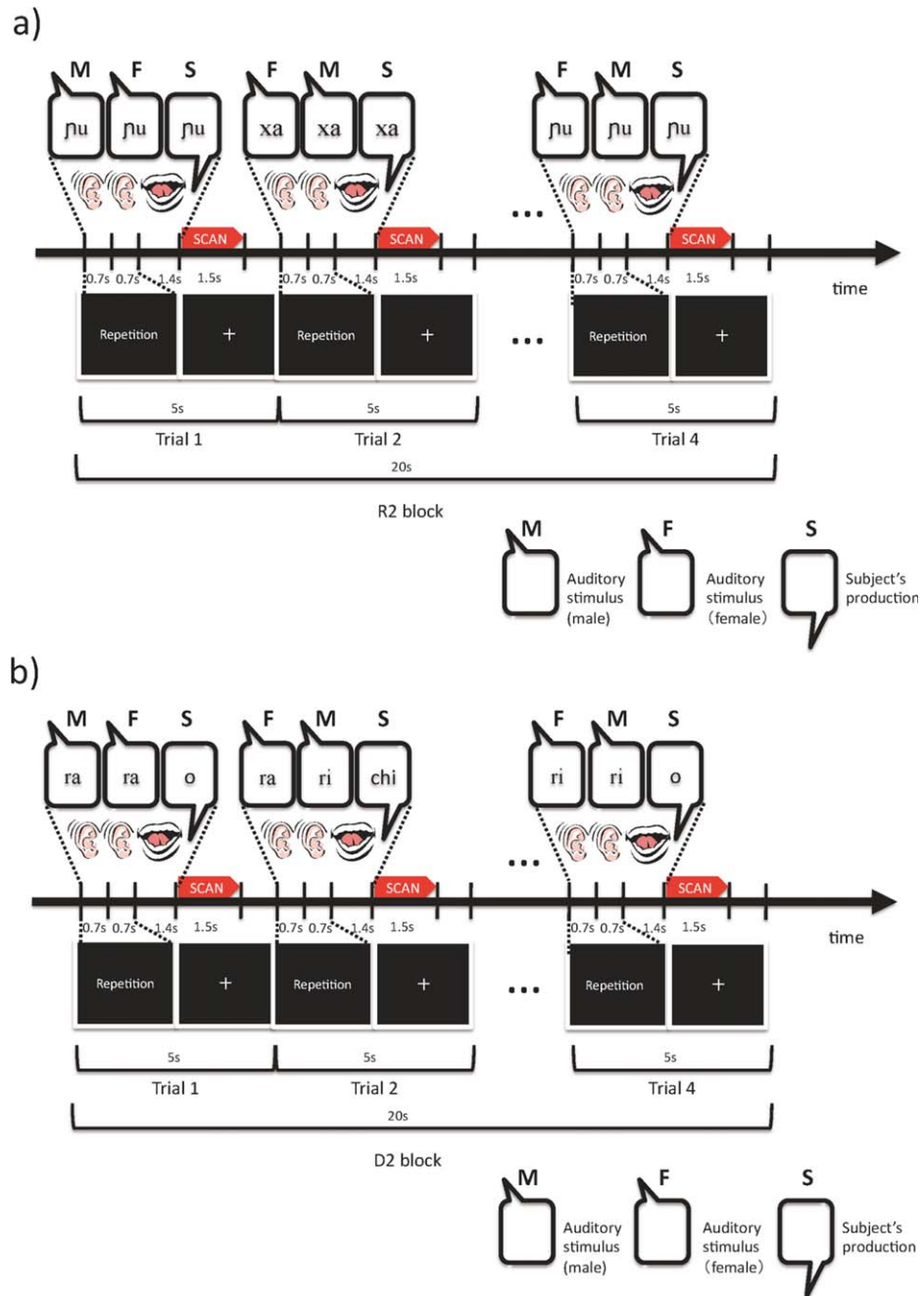
Figure 1.

Time courses of (**a**) the repetition tasks and (**b**) the discrimination tasks. In the repetition task, two of the same syllables spoken by different speakers (one female and one male) were presented sequentially in a trial. Participants were asked to overtly repeat the heard syllable once, reproducing the sound's phonological features as much as possible. In the discrimination task, two of same or different syllables spoken by different speakers were presented sequentially in a trial. Participants were asked to judge whether the pair was phonetically the same with a predetermined overt response of "O" or "Chi." See the Materials and Methods for further details. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

images (EPI) (echo time = 30 ms, flip angle = 90°, slice thickness = 4 mm, slice gap = 1.3 mm, field of view = 192 mm, matrix size = 64 × 64, voxel size = 3 × 3 ×5.3 mm) covering the entire cerebrum were acquired at a repetition time of 5,000 ms, and 119 EPI volumes were acquired for each participant. To avoid scanning noise interference during the perception of audio stimuli and head motion-related artifacts during utterances, we utilized a sparse scanning technique (acquisition time = 1.5 sec). It would have been quite difficult for participants to clearly evaluate unfamiliar phonetic features during image acquisition. The following preprocessing procedures were performed using Statistical Parametric Mapping (SPM5) software (Wellcome Department of Imaging Neuroscience, London, UK) and MATLAB (MathWorks, Natick, MA): adjustment of acquisition-timing across slices, head-motion correction, spatial normalization using the MNI EPI template, and smoothing using a Gaussian kernel with a full-width at half-maximum of 6 mm. Imaging data of subjects that showed excessive motion (greater than 3° of rotation or greater than 3 mm of translation) within the scanner was excluded from statistical analyses.

Evaluation of Participants' Utterances

Two native speakers (one male and one female) for each language and speech sound evaluated the participants' performance in the L2 repetition task. Raters listened to recorded utterances through a pair of headphones and rated each utterance for degree of foreign accent by pushing one of nine buttons representing a scale ranging from 1 (very strong foreign accent) to 9 (no foreign accent). Evaluation procedures followed by previous foreign-accent studies have been reviewed by Piske et al. [2001]. The evaluation was conducted sound-by-sound. Raters listened to and judged each participant's utterances for a certain sound and then repeated this process twice; the participants' utterances were presented in a random order for each set. The first judgment allowed raters to adapt themselves to the evaluation procedures and to determine the extent of variation in participants' utterances. Thus, the scores in the first judgment were not used in the analyses. An average rating, based on the final two judgments of the two raters (the male rater and female raters) was obtained for each utterance. Individual scores for production performance (L2 production scores) were computed by averaging the rates across speech sounds for individual participants.

Statistical Analyses

A conventional two-level approach for block-design fMRI data was adopted using SPM5. Voxel-by-voxel multiple regression analyses of the expected signal changes under each condition (R2, R1, D2, and D1) were constructed using the hemodynamic response function provided by SPM5 and applied to the preprocessed images for each participant. In addition, to remove the effect of high-motion volumes, we adopted a similar method to the "scrubbing" approach [Power et al., 2012]. We calculated the frame-wise displacement (FD) [Power et al., 2012] as an index of head movement. Here, we defined the high-motion frame as FD > 1.0. To adopt this approach into fMRI analysis by SPM, we used the binarized FD (0 or 1) as covariates in the first-level statistical model, instead of physically removing the high-motion frame. Motion regressors computed from SPM realignment procedure was not adopted in the model in this study.

Second-level random effects statistical inference on contrasts of the parameter estimates was performed using a one-sample t test. To identify the main effect of the task, the brain activations during the tasks were contrasted, (R2 + R1) − (D2 + D1) (i.e., greater activation during the repetition task compared with the perceptual task, regardless of whether L2 or L1 was involved). To identify the main effect of language, the brain activations for each language were contrasted, (R2 + D2) − (R1 + D1) (i.e., greater activation in L2 compared with L1, regardless of repetition or discrimination). To identify the interaction between task and language, the contrast (R2 − R1) − (D2 − D1) was tested (i.e., greater activation during repetition in L2 than in L1 while controlling for perceptual difficulty). To identify the region in which activation during the production of unfamiliar speech sounds [i.e. (R2 − R1) − (D2 − D1)] was correlated with age, multiple regression analyses were conducted, adjusting the performance of tasks (L2 production and perceptual task scores for L2 and L1), as covariates, to rule out the possibility that any age effect on activation derived simply from performance. For the exploratory whole-brain analyses, the statistical threshold was set at $P < 0.001$ for height and corrected to $P < 0.05$ for multiple comparisons using cluster size, assuming the entire brain as a search volume.

Second, to identify age-related changes in activation in the left IFG pOp and aIns, which were previously found to be activated during production of L2 speech sounds by adults [Moser et al., 2009], we used a small volume correction (SVC) [Friston, 1997; Worsley et al., 1996] for those regions. Small volumes were defined according to the anatomical automatic labeling (AAL) [Tzourio-Mazoyer et al., 2002]. AAL predefined the left IFG pOp and insula but not the anterior part of the insula. Thus, we followed the method developed by Moser et al. [2009], in which they limited AAL insular volume to the anterior part by combining the Jerne database (http://hendrix.ei.dtu.dk/services/jerne/ninf/voi.html) (See Moser et al. [2009] for details). Focusing on these regions of interest (ROIs), we performed one-sample t tests (the main effects of task and language, and the interaction) and multiple regression analyses, adjusting the performance of tasks as covariates, to examine the correlation between brain activation and age. It is reasonable to apply anatomical ROIs based on studies of adults to children because no significant

differences in the accuracy of anatomical normalization exist between adults and children of the age of our sample [Burgund et al., 2002; Muzik et al., 2000]. In addition to the *a priori* ROI (the left IFG pOp and aIns), we also tested for age-related changes in brain activation during the repetition task relative to the perceptual task in the brain region (the bilateral sensorimotor area) found to be significantly activated during the repetition compared with the perceptual task according our whole-brain analyses. Functional ROIs were defined by the clusters of activation (R2 + R1) − (D2 + D1) shown by the statistical threshold of $P < 0.001$ and cluster size corrections for multiple comparisons, $P < 0.05$. In all the aforementioned ROI analyses, the statistical threshold was set at $P < 0.001$ for height and corrected to $P < 0.05$ for multiple comparisons using cluster size with SVC. Finally, as a *post hoc* analysis, we tested the correlation between age and brain activation in the contrast of (R2 − R1) − (D2 − D1) as well as R2–R1 and D2–D1 ($P < 0.05$) at the peak of the cluster (the left IFG pOp) obtained from the small-volume-corrected ROI analysis. Brain activation was averaged within the sphere of a 5-mm radius at the peak using the MarsBar ROI tools [Brett et al., 2002]. Repeated measures of ANCOVA were used to examine the interaction effect of ROI (the left IFG pOp and aIns) and age. Here, the center ($x = -34.2$, $y = 15.9$, $z = 1.68$) of mass of the ROI for the left aIns was used as the center of the sphere of a 5-mm radius.

RESULTS

Behavior

The overt responses recorded during the scan confirmed that all participants executed all tasks according to the instructions. In the repetition tasks (R2 and R1), all participants adequately repeated the speech sounds they heard with no omission errors. We excluded four participants (three out of those four subjects from the age range 7–9 years and one from the age range 11–13 years) who performed the discrimination tasks (D2 and D1) poorly (accuracy rate for either D2 or D1 <80%) were excluded from statistical analyses to increase the number of successful trials within blocks. The average of accuracy in D2 and D1 were 90.7 ± 6.4% and 95.9 ± 5.1%, respectively. The maximum and minimum accuracy scores were 100% and 81.3%, respectively, in both D2 and D1. The average production score in R2 (range 1–9) was 4.8 ± 0.8. The maximum and minimum production scores in R2 were 3.1 and 6.3, respectively. The production score positively correlated with age ($r = 0.48$, $P = 0.003$; Table IV, Fig. 2). To check the influence of high leverage points, we also computed the correlation between the production score and age after excluding one subject from the 17–19 range, and two from the 15–17 range. The correlation was still significant ($r = 0.403$, $P = 0.018$). Additionally, we found a weaker but significant positive correlation between
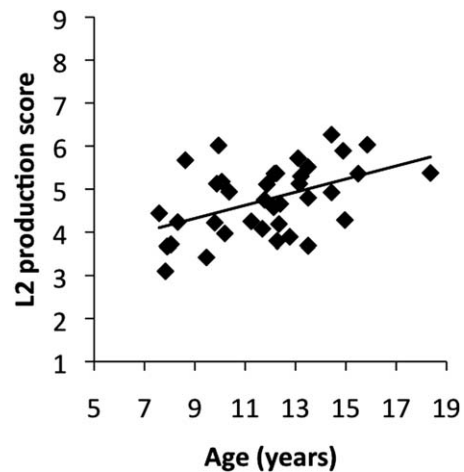

Figure 2.
Relationship between the production score for L2 speech sounds and age ($r = 0.48$, $P = 0.003$). A line fitted by linear regression is also shown.

production score and D2 ($r = 0.34$, $P = 0.04$) as well as between age and D1 score ($r = 0.36$, $P = 0.03$) (Table IV).

Imaging

Brain areas showing significant activation during the repetition compared with the perceptual task according to the whole-brain analyses are shown in Table V and Figure 3. The mouth regions in the bilateral primary sensorimotor areas were activated more significantly during the repetition task relative to the perceptual task, regardless of input stimulus (L2 or L1). However, no region showed significant activation or either a main effect of language or a task × language interaction. Voxel-based whole-brain linear regression analyses showed no significant correlation between age and the main effects and interaction.

In the small-volume-corrected ROI analyses, we found a significant positive correlation between the brain activation in the left IFG pOp associated with L2 speech sound production in the contrast (R2 − R1) − (D2 − D1) and age (Fig. 4, Table VI). The mouth region of the bilateral primary sensorimotor area, which was implicated in the main effect of repetition, did not show a significant correlation

TABLE IV. Correlation coefficients for L2 production score, D2 score, D1 score, and age

r (P)	D2 score	D1 score	Age
L2 production score	0.34[a] (0.04)	0.06 (0.71)	0.48[a] (0.003)
D2 score	–	−0.12 (0.47)	−0.05 (0.78)
D1 score	–	–	0.36[a] (0.03)

[a]$P < 0.05$

TABLE V. Local maxima of brain activation in the repetition compared with the discrimination condition

Area		MNI coordinate (x, y, z)	t-value	Cluster size (ke)
Pre/postcentral gyrus	L	−48, −6, 24	4.40	46
(sensorimotor area)[a]	R	54, −9, 30	4.97	55

[a]$P < 0.001$ for height and corrected to $P < 0.05$ for multiple comparisons using cluster size

between age and activation during the repetition task relative to the perceptual task.

In the *post hoc* analyses, we found a significant positive correlation between age and activation in the left IFG pOp in the contrast R2–R1 ($r = 0.37$, $P = 0.025$; Fig. 4c), but not in the contrast D2–D1 ($r = −0.12$, $P = 0.469$; Fig. 4d). This result confirms that the positive correlation between age and the interaction contrast in the left IFG pOp that was found in the small-volume-corrected ROI analysis was attributable to the activation increase for R2 with age, not D1. To confirm that the possible correlation between head movement and age did not significantly affect our results, we computed the session-averaged FD for each subject. There were no significant correlations between the averaged FD and age ($r = −0.26$, $P = 0.12$). Repeated measures of ANCOVA showed a significant main effect of age ($F(1, 35) = 4.65$, $P = 0.038$) as well as a significant interaction effect of ROI (the left IFG pOp and aIns) and age ($F(1,35) = 9.60$, $P = 0.004$).

DISCUSSION

This is the first study to investigate the developmental changes in brain activation involved in the production of L2 speech sounds by children. Consistent with our hypothesis, activation in the L-IFG pOp during the pro-
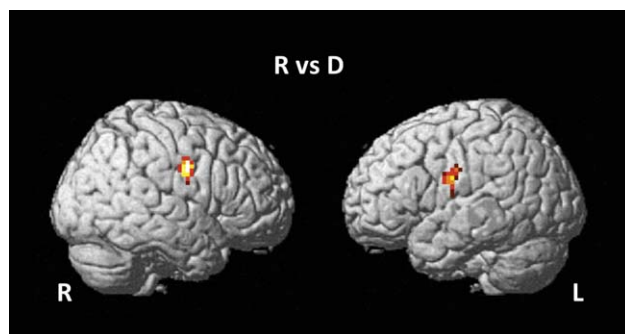

Figure 3.
Brain areas showing brain activation during a repetition task compared with a discrimination task. Activation in the bilateral sensorimotor areas was significantly greater in the repetition compared to the discrimination tasks.

duction of L2 speech sounds increased with age even after adjusting for the performance of tasks as covariates. The ability to produce L2 speech sounds (L2 production score) also increased with age. Our results suggest that the left IFG pOp, which is thought to play a role in speech motor planning in adults [Bohland and Guenther, 2006; Guenther et al., 2006; Papoutsi et al., 2009; Riecker et al., 2008; Soros et al., 2006], become more recruited during the production of L2 speech sounds with age, possibly due to the developmental specialization of brain function [Bitan et al., 2007; Brown et al., 2005; Schapiro et al., 2004; Tamm et al., 2002] related to speech motor planning.

The difference between the left IFG pOp and the left aIns in the age-dependency of activation during the production of L2 speech sounds provides some insight into the neural mechanisms underlying improvements in children's ability to produce L2 speech. In our study, the left IFG pOp showed a positive correlation between activation and age, but the left aIns showed no significant correlation. Moser et al. [2009] suggested that the left aIns facilitates the production of new motor plans for speech and the IFG pOp orchestrates the serial organization of established motor plans. Thus, our results may suggest that older children may combine established motor plans (i.e., L1 sounds) promptly and skillfully as a substitute for producing L2 speech sounds, whereas younger children may not be able to do this, which may result in an advantage for the older children with respect to immediate imitative L2 speech production [e.g., Snow and Hoefnagel-Höhle, 1977]. On the other hand, a series of behavioral foreign-accent studies have shown that children who immigrated early had less pronounced foreign accents, supporting the idea that younger children have an advantage over older children in learning pronunciation [Piske et al., 2001 for review]. Additionally, behavioral longitudinal studies have shown that older children do indeed have an initial advantage compared with younger children but that younger children eventually surpass older ones, who level off at a lower level [Aoyama et al., 2004; Oh et al., 2011; Snow and Hoefnagel-Höhle, 1977]. Thus, the ability of older children to promptly and skillfully incorporate a motor plan into their existing repertoire (i.e., L1), which is suggested in this study, may not establish the L2 sound system but may instead interfere with it. According to the Speech Learning Model [Flege, 1995], younger learners are more likely to establish new L2 speech sounds. This proposal is based on the hypothesis that as the phonetic units in the L1 sound system develop, they become greater attractors of L2 speech sounds, thereby suppressing the establishment of new phonetic categories for L2 speech sounds [Aoyama et al., 2004]. This model does not contradict our proposal.

The age-related increased activation during L2 speech production may be explained by the developmental specialization not only of speech motor planning but also of the human mirror neuron system (MNS) during childhood. The MNS is thought to be generally involved in the
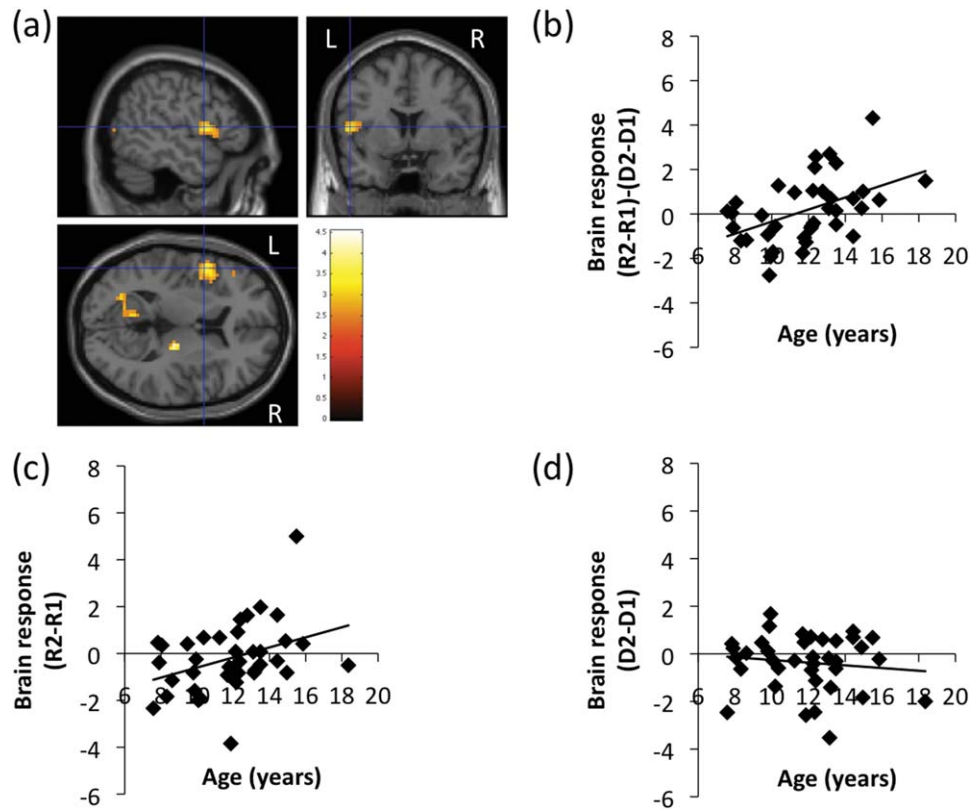
Figure 4.

(a) Increased brain activation in the left inferior frontal gyrus (IFG) pars oparcularis (pOp) during speech sound production in L2 compared with L1 [i.e. (R2 − R1) − (D2 − D1)] as a function of age via regression analyses. A liberal statistical threshold was applied for visualization purposes ($P < 0.01$, uncorrected for multiple comparisons). The peak of the cluster was $P <$ 0.001 for height and corrected to $P < 0.05$ for multiple comparisons using cluster size with SVC. Relationship of the brain activation in the left IFG pOp for (b) (R2 − R1) − (D2 − D1) ($r = 0.47$, $P = 0.003$), (c) R2–R1 ($r = 0.37$, $P = 0.025$), (d) D2–D1 ($r = -0.12$, $P = 0.469$) as a function of age. A line fitted by linear regression is also plotted in (b), (c), and (d).

imitation of familiar [Iacoboni, 2009; Iacoboni and Dapretto, 2006; Iacoboni et al., 1999; Rizzolatti and Craighero, 2004] and, especially, unfamiliar/novel actions [Buccino et al., 2004; Vogt et al., 2007]. The left IFG pOp, which is associated with speech motor planning, is also one of the core MNS circuits in humans [Iacoboni, 2009; Iacoboni and

Dapretto, 2006; Iacoboni et al., 1999; Rizzolatti and Craighero, 2004]. Although it is still unclear how the human MNS develops, it has been hypothesized that this system continues to develop throughout childhood [Kilner and Blakemore, 2007]. An attractive possibility, therefore, is that ongoing MNS development, as suggested by the age-related increased activity in the left IFG pOp, results in the age-related improvement in unfamiliar speech–sound imitation. It has already been suggested that the MNS may be the precursor of language [Rizzolatti and Arbib, 1998], and several studies mention the association between speech motor planning and the MNS [Guenther et al., 2006], further supporting this hypothesis.

There was no significant correlation between age and the activation during the repetition task relative to the perceptual task in the mouth regions of the bilateral primary sensorimotor area, which was activated during the repetition task relative to the perceptual task (Fig. 3). This may reflect earlier maturation of brain networks subserving

TABLE VI. Local maxima of brain regions with significant positive correlations between brain activation for (R2 − R1) − (D2 − D1) and age

		Z score	MNI coordinates (x, y, z) (mm)
Inferior frontal gyrus/pars opercularis[a]	L	3.31	−51, 6, 9

[a] $P < 0.001$ for height and corrected to $P < 0.05$ for multiple comparisons using cluster size with SVC

primary sensorimotor functions for speech production relative to L2 speech production, which is supported by the anatomical studies showing that anatomical regions underlying primary sensorimotor functions mature first, whereas higher-order association areas such as the prefrontal cortex mature later [Gogtay et al., 2004; Sowell et al., 2004]. Booth et al. [2003] also reported that the brain activation involved in relatively early matured function showed smaller developmental difference than relatively late-matured function, supporting our interpretation. Late maturation of the ability to produce L2 speech sounds may be reflected in the gradually increased activation in the left IFG pOp during the production of L2 speech sounds.

This study has at least one limitation. We mainly focused on the involvement of the left IFG pOp and aIns with the speech motor process during L2 speech production. However, the other processes, such as the perceptual ability to extract the detailed phonetic features of novel speech sounds, may also contribute to the L2 production ability. Although we included the scores of the perceptual discrimination task into the regression analysis as covariates, it was still possible that those scores did not adequately represent all of the perceptual abilities necessary for L2 speech production, which may have blurred the relationships between the L2 production score and brain activation.

In conclusion, we found that brain activation in the left IFG pOp during the production of L2 speech sounds as well as the accuracy of such L2 speech sounds increased with age. Our current findings suggest that the recruitment of the left IFG pOp during the production of L2 speech sounds was possibly enhanced due to the maturation of the neuronal circuits needed for speech motor planning. This, in turn, would lead to improvement in the ability to immediately imitate L2 speech. These data may help clarify the mechanism underlying age-related differences in learning a second language.

ACKNOWLEDGMENTS

REFERENCES

Aoyama K, Flege J, Guion S, Akahane-Yamada R, Yamada T (2004): Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. J Phon 32:233–250.

Bitan T, Cheon J, Lu D, Burman D, Gitelman D, Mesulam M, Booth J (2007): Developmental changes in activation and effective connectivity in phonological processing. Neuroimage 38: 564–575.

Bohland J, Guenther F (2006): An fMRI investigation of syllable sequence production. Neuroimage 32:821–841.

Booth J, Burman D, Meyer J, Lei Z, Trommer B, Davenport N, Li W, Parrish T, Gitelman D, Mesulam M (2003): Neural development of selective attention and response inhibition. Neuroimage 20:737–751.

Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL (2005): Developmental changes in human cerebral functional organization for word generation. Cereb Cortex 15:275–290.

Bunge S, Dudukovic N, Thomason M, Vaidya C, Gabrieli J (2002): Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. Neuron 33:301–311.

Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, Petersen SE, Schlaggar BL (2002): The feasibility of a common stereotactic space for children and adults in fMRI studies of development. Neuroimage 17:184–200.

Dronkers N (1996): A new brain region for coordinating speech articulation. Nature 384:159–161.

Flege J (1995): Second-language speech learning: Theory, findings, and problems. In Strange W, editor. Speech Perception and Linguistic Experience: Theoretical and Methodological Issues. Timonium, MD: York Press. pp 233–277.

Friston KJ (1997): Testing for anatomically specified regional effects. Hum Brain Mapp 5:133–136.

García Mayo M, García Lecumberri M (2003): Age and the acquisition of English as a foreign language. Clevedon, England: Multilingual Matters.

Gogtay N, Giedd J, Lusk L, Hayashi K, Greenstein D, Vaituzis A, Nugent T, Herman D, Clasen L, Toga A, Rapoport J, Thompson P (2004): Dynamic mapping of human cortical development during childhood through early adulthood. Proc Natl Acad Sci USA 101:8174–8179.

Guenther F, Ghosh S, Tourville J (2006): Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96:280–301.

Hillis A, Work M, Barker P, Jacobs M, Breese E, Maurer K (2004): Re-examining the brain regions crucial for orchestrating speech articulation. Brain 127:1479–1487.

Iacoboni M (2009): Neurobiology of imitation. Curr Opin Neurobiol 19:661–665.

Iacoboni M, Dapretto M (2006): The mirror neuron system and the consequences of its dysfunction. Nat Rev Neurosci 7:942–951.

Iacoboni M, Woods R, Brass M, Bekkering H, Mazziotta J, Rizzolatti G (1999): Cortical mechanisms of human imitation. Science 286:2526–2528.

Kilner J, Blakemore S (2007): How does the mirror neuron system change during development? Dev Sci 10:524–526.

Kohler E, Keysers C, Umiltà M, Fogassi L, Gallese V, Rizzolatti G (2002): Hearing sounds, understanding actions: Action representation in mirror neurons. Science 297:846–848.

Lenneberg E (1967): Biological foundations of language. Oxford, England: Wiley.

Mohr J, Pessin M, Finkelstein S, Funkenstein H, Duncan G, Davis K (1978): Broca aphasia: Pathologic and clinical. Neurology 28: 311–324.

Moser D, Fridriksson J, Bonilha L, Healy E, Baylis G, Baker J, Rorden C (2009): Neural recruitment for the production of native and novel speech sounds. Neuroimage 46:549–557.

Muzik O, Chugani DC, Juhasz C, Shen CG, Chugani HT (2000): Statistical parametric mapping: Assessment of application in children. Neuroimage 12:538–549.

Nagao M, Takeda K, Komori T, Isozaki E, Hirai S (1999): Apraxia of speech associated with an infarct in the precentral gyrus of the insula. Neuroradiology 41:356–357.

Ogar J, Willock S, Baldo J, Wilkins D, Ludy C, Dronkers N (2006): Clinical and anatomical correlates of apraxia of speech. Brain Lang 97:343–350.

Oh G, Guion-Anderson S, Aoyama K, Flege J, Akahane-Yamada R, Yamada T (2011): A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. J Phon 39:156–167.

Oldfield R (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Olson L, Samuels S (1973): Relationship between age and accuracy of foreign language pronunciation. J Educ Res 66:263–268.

Papoutsi M, de Zwart J, Jansma J, Pickering M, Bednar J, Horwitz B (2009): From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. Cereb Cortex 19:2156–2165.

Piske T, MacKay I, Flege J (2001): Factors affecting degree of foreign accent in an L2: A review. J Phon 29:191–215.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59:2142–2154.

Riecker A, Brendel B, Ziegler W, Erb M, Ackermann H (2008): The influence of syllable onset complexity and syllable frequency on speech motor control. Brain Lang 107:102–13.

Rizzolatti G, Arbib M (1998): Language within our grasp. Trends Neurosci 21:188–194.

Rizzolatti G, Craighero L (2004): The mirror-neuron system. Ann Rev Neurosci 27:169–192.

Schapiro M, Schmithorst V, Wilke M, Byars A, Strawsburg R, Holland S (2004): BOLD fMRI signal increases with age in selected brain regions in children. Neuroreport 15:2575–2578.

Simmonds AJ, Wise RJS, Leech R (2011a): Two tongues, one brain: Imaging bilingual speech production. Front Psychol 2:166.

Simmonds AJ, Wise RJS, Dhanjal NS, Leech R (2011b): A comparison of sensory–motor activity during speech in first and second languages. J Neurophysiol 106:470–478.

Snow C, Hoefnagel-Höhle M (1977): Age differences in pronunciation of foreign sounds. Lang Speech 20:357–365.

Soros P, Sokoloff L, Bose A, McIntosh A, Graham S, Stuss D (2006): Clustered functional MRI of overt speech production. Neuroimage 32:376–387.

Sowell E, Thompson P, Leonard C, Welcome S, Kan E, Toga A (2004): Longitudinal mapping of cortical thickness and brain growth in normal children. J Neurosci 24:8223–8231.

Taki Y, Hashizume H, Sassa Y, Takeuchi H, Wu K, Asano M, Asano K, Fukuda H, Kawashima R (2011a): Correlation between gray matter density-adjusted brain perfusion and age using brain MR Images of 202 healthy children. Hum Brain Mapp 32:1973–1985.

Taki Y, Hashizume H, Sassa Y, Takeuchi H, Wu K, Asano M, Asano K, Fukuda H, Kawashima R (2011b): Gender differences in partial-volume corrected brain perfusion using brain MRI in healthy children. Neuroimage 58:709–715.

Tamm L, Menon V, Reiss A (2002): Maturation of brain function associated with response inhibition. J Am Acad Child Adolesc Psychiatry 41:1231–1238.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996): A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.