Published in final edited form as:

Int J Billing. 2016 ; 20(3): 231–253. doi:10.1177/1367006914552206.

Speech sound learning depends on individuals' ability, not just experience

Pilar Archila-Suerte,

Department of Psychology, University of Houston, USA

Ferenc Bunta, and

Department of Communication Sciences and Disorders, University of Houston, USA

Arturo E. Hernandez

Department of Psychology, University of Houston, USA

Abstract

Aims—The goal of this study was to investigate if phonetic experience with two languages facilitated the learning of novel speech sounds or if general perceptual abilities independent of bilingualism played a role in this learning.

Method—The underlying neural mechanisms involved in novel speech sound learning were observed in groups of English monolinguals ($n = 20$), early Spanish–English bilinguals ($n = 24$), and experimentally derived subgroups of individuals with advanced ability to learn novel speech sound contrasts (ALs, $n = 28$) and individuals with non-advanced ability to learn novel speech sound contrasts (non-ALs, $n = 16$). Subjects participated in four consecutive sessions of phonetic training in which they listened to novel speech sounds embedded in Hungarian pseudowords. Participants completed two fMRI sessions, one before training and another one after training. While in the scanner, participants passively listened to the speech stimuli presented during training. A repeated measures behavioral analysis and ANOVA for fMRI data were conducted to investigate learning after training.

Results and conclusions—The results showed that bilinguals did not significantly differ from monolinguals in the learning of novel sounds behaviorally. Instead, the behavioral results revealed that regardless of language group (monolingual or bilingual), ALs were better at discriminating pseudowords throughout the training than non-ALs. Neurally, region of interest (ROI) analysis showed increased activity in the superior temporal gyrus (STG) bilaterally in ALs relative to non-ALs after training. Bilinguals also showed greater STG activity than monolinguals. Extracted values from ROIs entered into a 2×2 MANOVA showed a main effect of performance, demonstrating that individual ability exerts a significant effect on learning novel speech sounds. In fact, advanced ability to learn novel speech sound contrasts appears to play a more significant role in speech sound learning than experience with two phonological systems.

Keywords

Bilingualism; speech; learning; phonology; fMRI

The overarching goal of this study was to test whether the successful learning of novel speech sounds would occur independent of bilingual experience or whether learning would be facilitated by bilingualism. Similar perceptual discrimination performance in monolinguals and bilinguals after training would suggest that bilinguals, regardless of phonetic experience with two languages, have comparable perceptual abilities to monolinguals and that monolinguals and bilinguals are just as likely to have advanced or non-advanced ability to learn novel speech sound contrasts. On the other hand, improved performance in bilinguals exclusively would suggest that previous phonetic experience has potentially rendered the perceptual system more flexible, thus promoting new learning. Perceptual flexibility, the enhanced ability to learn new speech sounds, may be the result of extended practice with speech sounds in bilingual individuals if they have more advanced novel speech sound learning skills than their monolingual peers. In order to test these competing hypotheses (bilingualism-independent perceptual learning abilities vs. bilingualism-dependent perceptual flexibility), English monolinguals and early Spanish–English bilinguals received phonetic training with novel Hungarian speech sounds embedded in pseudowords. Brain activity in response to the pseudowords was examined using functional MRI before and after training. All monolingual and bilingual subjects were also classified as either individuals with advanced ability to learn novel speech sound contrasts – advanced learners (ALs) or individuals with non-advanced ability to learn novel speech sound contrasts – non-advanced learners (non-ALs) based on their performance throughout the training to shed light on the premise of individual perceptual ability independent of bilingualism. We chose to use the terms individuals with advanced versus non-advanced ability to learn novel speech sound contrasts (ALs and non-ALs, respectively) rather than “good” versus “poor” learners or perceivers of speech, because we want to emphasize that all of our participants are within typical limits (i.e. have no known speech, language, or hearing disorders and are able to communicate in their language or languages). However, using these terms also allows us to emphasize that within the typical population, differences in speech perception skills exist that have been referred to in various terms, as reviewed in subsequent sections.

Studies of auditory training have shown that individuals, typically monolinguals, can be categorized as good or poor learners of lexical tones and second language phonemes after very limited exposure to the stimuli during training (Chandrasekaran, Sampath, & Wong, 2010; Gaab, Gaser, & Schlaug, 2006; Golestani & Zatorre, 2004; Wong, Perrachione, & Parrish, 2007). Two similar studies investigating perception of speech sounds in bilingual populations were also able to classify their subject pool into subgroups of good and poor learners (Diaz, Baus, Escera, Costa, & Sebastian-Galles, 2008; Sebastián-Gallés et al., 2012). Based on these studies, it appears that some individuals may have advanced perceptual abilities relative to others with similar background and language exposure. However, it has also been reported that professional musicians with extensive experience have enhanced auditory perceptual abilities relative to non-musicians (Pantev et al., 1998),

demonstrating that experience with a certain type of auditory input can enhance perceptual abilities. Taken together, these studies suggest two possible alternatives: (1) how well new speech sounds are learned may depend on the perceptual abilities of an individual *independent* of the amount of exposure to the stimuli, as demonstrated by studies of good and poor auditory learning; or (2) novel speech sound learning may depend on the amount of experience with diverse phonemes, as demonstrated by studies with musicians. In the area of speech processing, it is unclear how knowing two phonological systems potentially influences new speech sound learning, and it is also unknown to what extent individual ability affects novel speech sound learning.

Monolingual and bilingual speech learning

Behavioral studies of phonetic training have vastly demonstrated that learning new speech sounds is possible in adult monolinguals (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Pruitt, Jenkins, & Strange, 2005). Much of this learning has been attributed to the reassignment of weights to certain acoustic cues in the speech signal through selective attention (Flege, 2003; Iverson et al., 2003; Nitttrouer, Crowther, & Miller, 1998). Structural neuroimaging studies have reported that individual differences in novel speech learning in monolinguals are related to brain structure; for instance, faster phonetic learners show greater white matter volumes in left parietal regions than slower learners (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Golestani, Paus, & Zatorre, 2002; Golestani, Price, & Scott, 2011; Golestani & Zatorre, 2009). Functional neuroimaging studies have also shown increased activation in bilateral frontal and temporal regions associated with speech processing after phonetic training relative to before training (Callan, Tajima, Callan, Akahane-Yamada, & Masaki, 2001). Difficult unfamiliar speech sound contrasts, for example, generate bilateral activity in the inferior frontal gyrus (IFG), superior temporal gyrus (STG), and supramarginal gyrus (Callan et al., 2003). Similarly, a pre-post MEG study looking at the perceptual improvement of the English /l/-/r/ distinction in Japanese monolingual listeners found greater neural sensitivity in the left-hemisphere mismatch field – the hemisphere commonly associated with language processing (Zhang et al., 2009). The results from these phonetic training studies are consistent with the speech perception literature, which repeatedly shows the left posterior-superior temporal gyrus (p-STG) to be involved in speech processing (Binder, 2000).

The behavioral literature on novel speech sound acquisition in bilinguals (i.e. bilinguals exposed to a *third* phonological system with which they are unfamiliar) has yielded mixed results. Some studies have suggested that phonetic experience with two or more languages enhances perceptual flexibility. For example, it has been found that Greek-English bilinguals discriminate and produce Thai /ba/ and /pa/ tokens better than English monolinguals (Beach, Burnham, & Kitamura, 2001), that bilinguals have an advantage when repeating novel phonemic sequences in comparison to monolinguals (Cohen, Tucker, & Lambert, 1967), and that multilinguals identify and discriminate single and gemminate stops in Japanese (e.g. “iken” vs. “ikken”) significantly more accurately than monolinguals (Enomoto, 1994). However, other studies indicate that phonetic expertise does not promote perceptual flexibility; therefore, perception of novel speech is not enhanced by previous phonetic experience (Pallier, Bosch, & Sebastian-Galles, 1997; Werker, 1986). It appears

that bilinguals can only improve the perception of second language sounds when training with the same second language phonemes subjects are already learning outside the laboratory; as in the case of Japanese–English second language learners receiving phonetic training to differentiate English /l/ and /r/ (Zhang et al., 2000). Therefore, it has to be noted that some of the studies cited above may only test novel aspects of existing speech contrasts (such as duration as in the case of geminates or voice onset time) rather than truly novel speech sounds. To date, research has not fully determined if bilinguals who have been exposed to two phonetic environments have a more malleable perceptual system than their monolingual peers that could enhance the ability of bilinguals to learn new speech sounds with which they are unfamiliar.

Recent neuroanatomical evidence has shown that bilinguals have larger bilateral Heschl's gyri than monolinguals (Ressel et al., 2012) and that bilinguals who are good perceivers of the second language have a thinner cortex in the left middle temporal gyrus and angular gyrus than bilinguals who are poor perceivers (Burgaleta, Baus, Diaz, & Sebastian-Galles, 2014). Although these studies tell us about how brain structures differ between monolinguals and bilinguals and how these structural differences relate to perceptual abilities *within* the group of bilinguals, these studies do not tell us how bilingualism might facilitate the perception of new speech sounds or how perceptual ability might be present independent of bilingualism. The present study is unique in that it looks at perceptual abilities in speech sound learning, taking bilingualism into consideration.

Even though neuroimaging phonetic training studies with bilinguals learning novel speech sounds have not been conducted, we hypothesize that if bilingual subjects behaviorally outperform monolinguals learning novel speech – such differences, if found, would demonstrate that experience with two phonological systems promotes learning of novel speech sounds – then bilinguals should show greater magnitude and intensity of activity in the STG bilaterally as region of interest (ROI) than monolinguals after training.

Advanced versus non-advanced learners of speech

As stated above, it has also been documented that limited exposure to auditory stimuli can lead to successful learning. Studies in other areas of auditory perception, such as pitch perception, have found that after training, successful pitch learners show streamlined activation in the left p-STG whereas less successful learners show diffused activity in the right STG, right IFG, and prefrontal/ medial frontal areas associated with working memory and attentional effort (Wong et al., 2007). Similar to these results, Gaab and colleagues (2006) found that strong learners of pitch had more activity in the left Heschl's gyrus, p-STG, and supramarginal gyrus after training than before training. On the other hand, studies of phonological training in children with reading difficulties have shown intense activation of the STG in successful learners and reduced neural activity in the STG plus distributed activation along frontal regions of the brain in unsuccessful learners (Blau et al., 2010; Simos et al., 2002).

Other areas of expertise such as piano playing and music reading also support the notion that neural efficiency is represented as focused activity in the primary brain region associated

with the task (Bengtsson et al., 2005; Stewart et al., 2003). Golestani and Zatorre (2004) showed that successful learners of a second language recruit the same areas typically involved in the processing of native phoneme contrasts, including the left STG, insula-frontal operculum, and IFG, when listening to non-native phonemes. In summary, the left STG appears to be the locus of plasticity in those who have improved their ability to make phonetic distinctions (Callan et al., 2003; Golestani & Zatorre, 2004; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006).

Based on the literature available to date, we hypothesize that if general perceptual ability varies independently of language experience with two languages, bilinguals will not differ significantly from monolinguals in learning novel speech sounds after phonetic training if those sounds are equally novel for all participants. That is, bilingualism will not facilitate the learning of novel sounds, making ALs and non-ALs equally likely to be represented in the groups of monolinguals and bilinguals. If this is the case behaviorally, then regardless of language group (monolingual or bilingual), all individuals with advanced ability to learn novel speech sound contrasts (ALs) will demonstrate significantly more improvement in discriminating new speech sounds after training than individuals with non-advanced ability to learn novel speech sound contrasts (non-ALs). Neurally, ALs are expected to show increased activity in the ROI of the STG bilaterally, but primarily increased activity in the left hemisphere as it has been demonstrated in other studies of successful speech learning. On the other hand, non-ALs are expected to show weaker or diffused activity in the ROI of the STG bilaterally. This might be represented as smaller clusters or less intense peaks of activation, in line with previous findings of unsuccessful learning.

A secondary objective of the present study is to uncover if novel speech sounds from the same foreign phonemic category (e.g. /y/-/y/) or from a different phonemic category (e.g. /y/-/æ/) are learned and processed differently by good and poor learners and the respective language groups of monolinguals and bilinguals. According to the literature, phonemes from the same category, known as “within-category”, tend to be difficult to discriminate – for native speakers of the language – because the acoustic cues that characterize them are not relevant for differentiating phonemes in a given language (Best & McRoberts, 2003; Pisoni & Tash, 1974). In other words, the within-category differences that may exist do not distinguish phonemes in a given phonological system, because the differences between sounds within the same phoneme category are not linguistically relevant in the given language. Consequently, native speakers tend to treat allophonic variations in speech as sounds of the same category and non-phonemically distinct. On the other hand, phonemes that belong to two different categories, known as “between-category” differences, include acoustic cues that are relevant for the given phonological system with respect to differentiating phonemes (Pisoni & Tash, 1974; Wood, 1976).

The problem for non-native speakers of a language when acquiring novel speech sounds from an unfamiliar language is that all acoustic events act as potential cues to differentiating novel phonemes. Thus, it stands to reason that individuals who have a more advanced ability to learn new phonemic contrasts have a potential advantage in the acquisition of a new phonological system. Individuals who can tune into the *relevant* differences of novel speech stimuli may be able to acquire phonological systems better than people whose perceptual

ability is less advanced. One of the questions addressed in the present paper is whether bilingualism promotes the ability to tune into novel speech sound contrasts or if advanced learners of novel speech sound contrasts (ALs) are equally present in bilingual and monolingual individuals.

An fMRI study with monolinguals found that stimuli from different phonemic categories evoke activity in the posterior region of the left superior temporal sulcus (STS) and supramarginal gyrus, whereas stimuli from the same phonemic category evoke activity in a subcortical region of the left caudate nucleus (Joanisse, Zevin, & McCandliss, 2007). Given the restriction of the present fMRI analysis to the STG region to decrease type I error, ALs are expected to show increased activity in response to between-category and within-category stimuli in bilateral STG relative to non-ALs after training. If bilinguals' experience with two phonetic systems enhances their ability to discriminate and learn novel speech sounds from within-categories and between-categories, then bilinguals are expected to show as intense and widespread activity within the bilateral STG as ALs. Non-ALs are expected to be unable to behaviorally discriminate novel speech sounds; therefore, poor learners are expected to show reduced or diffused activity in the STG.

Method

Participants

Forty-four English monolinguals and early Spanish–English bilinguals (who were exposed to both of their languages before five years of age) participated in this study. There were 20 participants in the monolingual group (eight men, 12 women) between 19 and 27 years of age, and 24 participants in the bilingual group (seven men, 17 women) between 18 and 31 years of age. In the bilingual group, participants learned Spanish as their first language and *sequentially* learned English as their second language. On average, participants completed 16.6 years of education by the time of testing. All participants were right-handed according to the Edinburgh inventory (Oldfield, 1971), reported no history of speech or language disorders, and consented to the protocol approved by the committee for the protection of human subjects at the University of Houston.

Procedure

This study consisted of seven lab visits. During the first visit, participants completed a language history questionnaire that verified eligibility for the study. Participants' language proficiency was assessed using the Woodcock Language Proficiency Battery – Revised (Woodcock, 1995). At the end of the first visit, a trained lab assistant scheduled the participant's next lab visits in the following order: *day 2*: pre-training fMRI session, *day 3*: phonetic training session 1, *day 4*: phonetic training session 2, *day 5*: phonetic training session 3, *day 6*: phonetic training session 4, and *day 7*: post-training fMRI session. Before starting the first phonetic training session, participants completed a five-minute behavioral pretest. The pretest evaluation was used as a baseline score. Additionally, after each training session, participants completed a five-minute posttest for a total of four posttests. The study was concluded with the post-training fMRI session. During pretest, training sessions, and posttests, participants were asked to discriminate if the pseudowords heard were the same or

different. Using a button box, participants pressed a button to indicate “same” and another button to indicate “different” pairs. The presentation of each trial was randomized across participants. Each training session lasted approximately 25 minutes. Seventy-seven percent of the subjects initially enrolled successfully completed the entire study.

Behavioral measures

Online language history questionnaire—This survey collected demographic, medical, academic, socioeconomic, and linguistic background information.

Woodcock language proficiency battery – Revised (WLPB-R)—The tests of picture vocabulary and listening comprehension were selected to assess overall expressive and receptive abilities in English (Woodcock, 1995). Bilinguals also completed the tests in Spanish (Woodcock & Muñoz-Sandoval, 1995). Participants provided the label of different objects, animals, and professions for the picture vocabulary test and filled in the blank to complete sentences for the listening comprehension test.

Stimuli

Eight different Hungarian vowel sounds were used to create pseudowords for the phonetic training and fMRI tasks. The target sounds were: /œ:/, /ø:/, /u:/, /u/, /o:/, /y/, /y:/, /o/. These vowel sounds were selected because they either do not exist in English or Spanish or they have similar analogs in the inventories of both languages. See Figure 1 for a depiction of vowel distribution in English, Spanish, and Hungarian in the vowel chart of the International Phonetic Alphabet (IPA). Except for the final consonant, /d/, all initial and medial consonants (voiced and voiceless palatal stops, /j/ and /c/, respectively) were novel for both English and Spanish–English bilinguals. The stimuli were recorded in a sound-treated room using an Audio Technica microphone AT 4040 and an external audio card M-audio fast pro at the Research Institute for Linguistics of the Hungarian Academy of Sciences in Budapest, Hungary. The sampling rate of the mono recording was 44 kHz and 16-bit resolution. In order to prevent biasing the monolingual group to perform better with pseudowords that are more likely to fit their language repertoire (i.e. monosyllables), we used *disyllabic pseudowords* that are equally likely to be part of English or Spanish. The Hungarian consonants that constituted the pseudowords were /j/ and /c/, as noted above. These sounds do not have analogs in English or Spanish. The consonants /j/ and /c/ (written as “gy” and “ty” in Hungarian, respectively) remained in the same position (initial /j/ and medial /c/) across all pseudowords and only the vowels were manipulated. The Hungarian consonant /d/ was included in word-final position to resemble the typical ending of Spanish and English words. Once the recording of pseudowords was completed, these were paired to create the conditions of *same* and *different*. *Same* pairs contained the same pseudowords read by different speakers. *Different* pairs contained different pseudowords read by different speakers. Eight types of pairs were used in this study, four pairs of *same* (“gyötyöd” /j œcœd/ – “gyötyöd” /j œcœd/; “gy ty d” /j ø:cø:d/ – “gy ty d” /j ø:cø:d/; “gyütyüd” /j ycyd/ – “gyütyüd” /j ycyd/; and “gy ty d” /j y:cy:d/ – “gy ty d” /j y:cy:d/) and four pairs of *different* (“gyötyöd” /j œcœd/ – “gyotyod” /j ocod/; “gy ty d” /j ø:cø:d/ – “gyótyód” /j ø:co:d/; “gy ty d” /j ø:cø:d/ – “gy ty d” /j y:cy:d/; and “gy ty d” /j y:cy:d/ – “gyütyüd” /j ycyd/).

Stimuli recordings

Eight native Hungarian speakers (four males, four females) between the ages of 27 and 33 recorded the pseudowords for this study. On average, speakers had resided in Budapest for 18.7 years and had completed 19 years of education at the time of recording. Speakers were asked to read the pseudowords “gyötyöd” /j œcœd/, “gy ty d” /jœ:cœ:d/, “gyütyüd” /ju:cu:d/, “gyutyud” /jucud/, “gyótyód” /jo:co:d/, “gyütyüd” /jycyd/, “gy ty d” /jy:cy:d/, and “gyotyod” /jocod/ following three different sets of instructions. In the first set, they were asked to enunciate the pseudowords carefully, emphasizing each vowel’s characteristic (e.g. height, backness, and duration). In the second set of instructions, speakers were asked to read the pseudowords a bit faster but still enunciating carefully. Finally, in the third set of instructions, speakers were asked to read the pseudowords at the same rate as regular conversation. To prompt a conversational quality, speakers read the pseudowords in carrier sentences.

High variability phonetic training (HVPT)

The high variability phonetic training (HVPT) paradigm contained natural speech from multiple speakers in multiple phonetic environments. HVPT was chosen because it is known to result in long-term improvements for up to six months and to generalize learning from trained stimuli to untrained stimuli (Iverson, Hazan, & Bannister, 2005; McCandliss et al., 2002). PsyScope X Build 57 (Cohen et al., 2010) was used to develop the training task in which participants learned to discriminate novel vowel sounds (same *vs.* different) with the help of computerized feedback; a beep for a correct response and a buzz for an incorrect response.

During training, the speakers’ rate of articulation described above (slow, fast but enunciated, and conversational-rate) was manipulated in three separate blocks. The first block contained trials that were slowly and carefully enunciated, the second block contained trials that were articulated a bit more quickly, and the third block contained trials that were articulated at the rate of standard conversation. Each trial consisted of two pseudowords spoken by different speakers of the same gender. Trials that presented the same pseudoword (e.g. gyötyöd /j œcœd/ – gyötyöd /j œcœd/) were referred to as *same* and trials that presented two different pseudowords (e.g. gy ty d /jœ:cœ:d/ – gy ty d /jy:cy:d/) were referred to as *different*. Therefore, one pair of pseudowords was equivalent to one trial. Within each block of training, there were 144 trials (48 trials of *same* and 96 trials of *different*), for a total of 432 trials. Two shorter versions of the training task without feedback were used as pretest and posttests. During training, pretest, and posttests, participants were asked to judge whether the pairs of pseudowords were the same or different by pressing the assigned button on the button box.

Neuroimaging measure

A pre-attentive listening paradigm was implemented for the pre-training and post-training fMRI tasks. Participants heard the Hungarian pseudowords through a pair of MRI-compatible headphones while a non-captioned muted movie, *Planet Earth*, displayed scenic views on the scanner’s projector screen. Participants were asked to attend to the movie while the pseudowords played in the background. No overt responses were collected. Pre-attentive

listening was used to create the environmental conditions most listeners are exposed to in their day-to-day experience and to prompt the automatic neural response of the auditory system that better maps onto the perceptual processes observed behaviorally (Joanisse et al., 2007).

A clustered volume acquisition design was employed to present the pairs of pseudowords (i.e. one trial) during an interval of silence between brain scans. This enabled participants to hear the pseudowords clearly without scanner noise. Each experimental trial lasted four seconds, including the 1.5 seconds of scanning time. The fMRI tasks also included baseline trials of silence. Five experimental trials plus three baseline trials in a row composed a block of stimuli that was either *same* or *different*. See Archila-Suerte, Zevin, Ramos, and Hernandez (2013) for an illustration of a trial and a block in the pre-attentive listening fMRI tasks. Each fMRI task lasted 34 minutes and contained 512 trials (64 blocks). To prevent a practice effect, the pre-training and post-training fMRI tasks presented slightly different types of trials.

AL versus non-AL groups—A hierarchical clustering analysis was conducted to define all the possible clusters from the dataset. The results indicated the presence of two clusters (see results section for details about the analysis). This was followed by *k*-means clustering, with fixed seeds ($k = 2$), which enabled the actual formation of clusters and the classification of participants into two groups respectively labeled as ALs and non-ALs depending on whether their performance-throughout-the-training mean score fell below or above the cutoff indicated in the clustering analysis.

As opposed to using the last posttest score to reflect final learning after training, high within-subject variability between posttest scores demonstrated that it was more appropriate to calculate an average score of all posttests for each participant. Participants who tended to score higher throughout the training received a higher overall score and participants who tended to score lower throughout the training received a lower overall score. This prevented misclassifying participants as ALs or non-ALs based on just the last posttest score.

fMRI acquisition parameters

Whole-brain scans were performed with a 3.0 Tesla Magnetom Trio (Siemens, Germany) at Baylor College of Medicine's Human Neuroimaging Laboratory in Houston, Texas. A total of 517 functional (T2*-weighted) images were acquired for each fMRI session (pre- and post-) using clustered volume acquisition (CVA) 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) = 4s, TR delay (silent interval) = 2.4s, volume acquisition time (TA) = 1.6s, transversal slices per volume = 26, TE = 30ms; 5 mm thickness, $3.4 \times 3.4 \times 5.0$ mm resolution, flip angle = 90 degrees, with the centermost slice aligned with the anterior commissure and posterior commissure (AC-PC). High-resolution anatomical images were collected with a T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (TR= 1.2s, TE= 2.66 ms, 1mm^3 isotropic voxel size) reconstructed into 192 slices. Auditory stimuli were presented using PsyScope. This software synchronized the task with the scanner with millisecond accuracy.

fMRI data analysis

Whole-brain and ROI analyses were conducted with SPM8 (Wellcome Trust Center for Neuroimaging, London, 2001) using a block design specification for the statistical model. Functional images were slice-time corrected, motion-corrected, aligned to anatomical scans, and normalized to Montreal Neurological Institute (MNI) stereotaxic space. Spatial smoothing used an 8 mm full-width half maximum Gaussian Kernel. At the first level of analysis, the stimulus timepoints of the pseudoword conditions were modeled with a canonical hemodynamic response function and contrasted across pre-training and post-training fMRI sessions for each participant. Four conditions made up the *same* category of novel speech and four conditions made up the *different* category of novel speech, as previously described. Lastly, a baseline condition consisting of dedicated periods of silence was included. All the *same* and *different* conditions were combined in order to create a new condition labeled *all novel speech* in order to investigate overall neural activity in response to novel sounds. This enabled us to generate a pairwise contrast (novel speech versus baseline) to compare these two main conditions between groups at the second level of analysis in two separate 2×2 ANOVA factorial designs: monolingual/bilingual × pre/post-training fMRI; and AL/non-AL × pre/post-training fMRI. The MNI coordinates and anatomical labeling were obtained through Anatomy Toolbox, a software extension to SPM (Eickhoff et al., 2005).

After the preliminary whole-brain analysis, two ROI were selected based on the literature in speech learning and applied in the 2×2 ANOVA described above. These regions were the right STG and left STG. The anatomical ROIs were created using WFU-PickAtlas (Maldjian, Laurienti, & Burdette, 2004; Maldjian, Laurienti, Kraft, & Burdette, 2003) and applied through the analysis pipeline in SPM. In the ROIs, the α value of < 0.05 was used for intensity threshold and $k > 20$ was used for the extent of cluster size. Unlike whole-brain analyses, which require family-wise-error (FWE) correction due to the voxel-by-voxel testing of the null hypothesis, ROI analysis conducts an average of the timeseries of all the voxels in the ROI, thereby allowing for contrasts with a probability of $p < 0.05$ to be considered statistically significant.

Results

Participants' characteristics

Monolingual ($n = 20$) and bilingual individuals ($n = 24$) did not differ in age $F_{(1,42)} = 3.02$, $p = 0.08$, years of education $F_{(1,42)} = 1.96$, $p = 0.16$, scholastic ability according to their self-reported grade point average (GPA) scores $F_{(1,42)} = 0.16$, $p = 0.68$, or receptive skills in English $F_{(1, 42)} = 2.54$, $p = 0.11$. The mean age of exposure to English of the bilingual participants was 3.61 years ($SD = 1.39$), thus conforming a group of early sequential bilinguals.

Behavioral results

Woodcock language proficiency battery – revised (WLPB-R)—Overall proficiency of English was obtained by combining the scores of picture vocabulary and listening comprehension, which were significantly correlated $r = 0.366$, $p = 0.015$ ($R^2 = 0.13$).

Monolinguals had a significantly higher English proficiency level than early bilinguals $F_{(1, 42)} = 5.173, p = 0.028; \eta_p^2 = 0.11$. Within the bilingual group, Spanish proficiency (also calculated by combining the scores of picture vocabulary and listening comprehension, $r = 0.746, p < 0.001; R^2 = 0.55$) was slightly lower ($M = 71.4, SD = 8.6$) than their English proficiency ($M = 76.1, SD = 9.09$). This indicates that bilingual participants were more proficient in English than in Spanish but not as proficient as their monolingual counterparts in English. See Figure 2 Panel A for a depiction of English proficiency in monolinguals and Spanish–English proficiency in bilinguals. When examining the overall linguistic ability of bilinguals by combining their word knowledge and understanding of both languages, however, bilinguals had significantly more expansive vocabularies and comprehension than monolinguals ($F_{(1, 42)} = 304.8, p < 0.001$).

Phonetic training—On average, monolinguals' performance from pretest to posttest 4 increased by 6.31% and bilinguals' performance increased by 5.95%. Although both groups significantly improved over time throughout the training $F_{(1, 41)} = 42.679, p < 0.001; \eta_p^2 = 0.510$, they did not differ significantly from each other at any point during the training $F_{(1, 41)} = 1.891, p = 0.177; \eta_p^2 = 0.44$. See Figure 2 Panel B. These findings demonstrate that phonetic experience with two languages does not appear to behaviorally enhance perceptual abilities in bilinguals.

To investigate whether general perceptual abilities play a role in novel speech sound learning regardless of language group, a mean score was calculated by combining the scores of each post-test. Because participants sometimes scored better on the third posttest than on the fourth posttest, it was more reliable to examine overall performance throughout the training than to only examine the score of the last posttest after training. See Figure 3(a) for a depiction of within-subject variability across posttests.

As explained in the method section, participants were experimentally classified as *ALs* or *non-ALs* based on hierarchical clustering and *k*-means procedure. Ward's cluster method and square Euclidean distance interval were employed in the hierarchical cluster analysis to determine minimum and maximum distances necessary for differentiating the groups, which must be relatively homogenous within themselves (minimum distance) and relatively heterogeneous compared to the other group (maximum distance). The distance coefficients in the agglomeration schedule exposed two clusters in the dataset (See Figure 3(b)). Participants who performed above the cutoff by correctly discriminating and therefore learning novel speech sounds were classified as *ALs* and participants who performed below the cutoff were classified as *non-ALs*. Here, *ALs* ($n = 28$; 14 monolinguals and 14 bilinguals) demonstrated a 9.86% average increase in learning after training, whereas *non-ALs* ($n = 16$; six monolinguals and 10 bilinguals) demonstrated a -0.42% average decrease in learning.

Similarly to the pattern of all novel speech sound learning, monolinguals and bilinguals did not significantly differ after training in the accuracy of identifying pseudowords that contained sounds from the same phonemic category or sounds that belonged to different phonemic categories (for *same* $F_{(1, 42)} = 0.191, p = 0.66; \eta_p^2 = .005$; for *different* $F_{(1, 42)} = 0.159, p = 0.69, \eta_p^2 = .004$). On the other hand, *ALs* identified pseudowords with speech

sounds from the same category significantly more accurately than non-ALs after training (for *same* $F_{(1,42)} = 38.78$, $p < 0.001$, $\eta_p^2 = .480$) but did not discriminate speech sounds from different categories significantly better than non-ALs ($F_{(1,42)} = 2.21$, $p = 0.145$, $\eta_p^2 = .05$). That is to say that according to the behavioral data, ALs were able to tune out non-phonemic differences more accurately than their non-AL peers, but differences were not (or not yet) evident in the cross-categorical (i.e. *different*) task. In this behavioral analysis, the raw scores of the four pseudoword pairs of *same* and the four pseudoword pairs of *different* were averaged out for each participant in order to have one score representing *same* and one score representing *different*.

Neuroimaging results

Whole-brain FWE-corrected within-group analysis: For the language grouping, bilinguals showed increased activity in the right STG before training ($k = 21$, $t = 4.76$) and bilateral STG after training (left $k = 679$, $t = 5.7$; right $k = 1084$, $t = 6.01$), whereas monolinguals only showed increased activity in the bilateral STG before training (left $k = 497$, $t = 6.19$; right $k = 552$, $t = 5.79$). For the experimentally derived groups of performance, ALs showed increased activity in the STG bilaterally before and after training (before: left $k = 995$, $t = 6.41$; right $k = 1006$, $t = 5.9$. After: left $k = 1232$, $t = 6.62$; right $k = 1152$, $t = 6.44$). ALs also showed additional activity in regions of the thalamus, caudate, and posterior cingulate gyrus after training. Non-ALs did not show any increased activity in this whole-brain analysis before or after training. To prevent an overextension of the manuscript, only the most relevant figures are presented. For all other details about intensity and size of brain activation in whole-brain and ROI analysis, see Table 1.

Region of interest analysis: In the ROI analyses below, the reverse comparison as well as the expected straight comparison, are reported as is pertinent to this study. Conducting the reverse comparison in each ROI serves as a confirmatory measure to demonstrate the robustness of the results for each group.

Bilinguals versus monolinguals: Bilinguals showed increased activity in the ROIs of the left and right STG after training relative to monolinguals in response to novel speech versus baseline (left STG $k = 278$, $t = 2.42$; right STG $k = 428$, $t = 2.33$). See Figure 4(a). Compared to bilinguals (in the reverse analysis), monolinguals did not show any activity in these ROIs after training.

ALs versus non-ALs: ALs showed increased activity in the ROIs of the left and right STG after training relative to non-ALs in response to novel speech versus baseline. For the ALs, the left STG showed two clusters (cluster 1 $k = 290$, $t = 2.84$; cluster 2 $k = 201$, $t = 2.75$) and the right STG showed three clusters (cluster 1 $k = 113$, $t = 2.55$; cluster 2 $k = 178$, $t = 2.28$; cluster 3 $k = 31$, $t = 2.23$). See Figure 4(a). Increased brain activity was not observed in either left or right STG ROIs for the reverse contrast of non-ALs versus ALs after training.

Contrasts within performance and across language

Bilingual ALs versus monolingual ALs: Bilinguals who were ALs showed increased activity in ROI of the right STG relative to monolinguals who were also ALs ($k = 331$, $t =$

2.45). The reverse contrast did not show more activity in the pre-selected ROIs for monolingual ALs compared to their bilingual AL peers.

Bilingual non-ALs versus monolingual non-ALs: Bilinguals who were non-ALs showed increased activity in ROIs of the right and left STG relative to monolinguals who were also non-ALs. There were two clusters in the ROI of the left STG (cluster 1 $k = 39$, $t = 2.05$; cluster 2 $k = 28$, $t = 2.03$) and two clusters in the ROI of the right STG (cluster 1 $k = 68$, $t = 2.23$; cluster 2 $k = 22$, $t = 2.11$). According to the reverse contrast, monolingual non-ALs did not show increased activity in the preselected ROIs compared to their bilingual non-AL peers.

Within-category and between-category comparisons

Bilinguals versus monolinguals: Bilinguals showed increased activity in the STG bilaterally in response to phonemes from within- and between-categories compared to monolinguals after training (within-category: left $k = 51$, $t = 2.19$; right $k = 247$, $t = 2.19$. Between-category: left $k = 786$, $t = 2.58$; right $k = 609$, $t = 2.53$). See Figure 4(b) for a depiction of activity related to between-category stimuli. Based on the reverse analysis, monolinguals did not show increased activity in response to either type of speech categories compared to bilinguals after training.

ALs versus non-ALs: Compared to non-ALs, ALs showed activity in three clusters of the left STG in response to phonemes from within-categories (cluster 1 $k = 127$, $t = 2.44$; cluster 2 $k = 23$, $t = 2.07$; cluster 3 $k = 44$, $t = 2.05$) and activity in bilateral STG in response to between-category stimuli after training. For between-category pairs, there were three clusters in the left STG (cluster 1 $k = 444$, $t = 3.36$; cluster 2 $k = 267$, $t = 2.87$; cluster 3 $k = 29$, $t = 2.15$) and two clusters in the right STG (cluster 1 $k = 302$, $t = 3.14$; cluster 2 $k = 523$, $t = 2.54$). See Figure 4(b). Regarding the reverse contrast, non-ALs did not show increased activity in response to either type of speech sound contrast categories compared to ALs after training.

To determine whether language group (bilingual versus monolingual) or performance (AL versus non-AL) had a greater effect in learning novel speech sounds, values of cluster magnitude were extracted from the right- and left-hemisphere STG ROIs for each participant and entered into a 2×2 MANOVA (bilingual/monolingual × AL/non-AL) in SPSS. This analysis showed a main effect of performance after training (i.e. post-training fMRI scan) for the magnitude of activation in the right STG ($F_{(1, 35)} = 5.16$, $p = 0.02$, $\eta_p^2 = 0.129$), and left STG ($F_{(1, 35)} = 4.20$, $p = 0.04$, $\eta_p^2 = 0.107$). There was no main effect of language group for the magnitude of activation in either the left hemisphere ($F_{(1, 35)} = 0.001$, $p = 0.971$) or right hemisphere ($F_{(1, 35)} = 0.001$, $p = 0.978$), and no performance by language group interaction in the left ($F_{(1, 35)} = 0.143$, $p = 0.708$) or right hemisphere ($F_{(1, 35)} = 0.016$, $p = 0.89$). This indicated that ALs had larger clusters of activation in the right and left STG after training relative to non-ALs. See Figures 5(a) and 5(b) for an illustration of the performance group main effect on cluster magnitude in the right and left hemisphere after training.

Discussion

The present study investigated whether perceptual abilities independent of bilingualism played a role in novel speech sound learning, or whether extended experience with two phonological systems facilitated this learning in adulthood. The results indicate that the capacity to learn new speech sounds in the context of pseudowords is more likely to result from a particular ability to discriminate speech cues than from experience with two languages. Even though bilinguals showed increased neural activity in the STG relative to monolinguals after training, it is difficult to conclude with certainty that bilingualism facilitates novel speech sound learning more than advanced perceptual ability because bilinguals – as a group – were not significantly better than monolinguals at discriminating pseudowords in the behavioral task.

Behaviorally, all participants demonstrated having significantly improved in learning after training. This indicated that the design of the phonetic training task – high variability in speech mixed with exaggerated cues – met the goal of teaching the pseudoword stimuli to naïve listeners. The results showed that bilinguals did not significantly differ from monolinguals in learning novel speech sounds. Instead, there was a significant difference between the experimentally derived groups of ALs and non-ALs when discriminating novel speech stimuli. This behavioral finding suggests that learning new speech sounds depends significantly on each individual's ability to perceive and discriminate novel speech sounds – an ability that appears to exist in both bilinguals and monolinguals. In the subset of monolingual ALs, the “lack” of experience with multiple linguistic systems did not impede them from successfully discriminating perceptually challenging pseudowords with novel speech sounds from a language with which they were unfamiliar. Furthermore, a person's *inability* to discriminate novel speech sound contrasts appears to be relatively unaffected by the amount of practice obtained through a phonetically enriched bilingual environment, as in the case of bilingual non-ALs. Bilingual non-ALs were presumably exposed to large amounts of diverse speech input, yet this experience did not assist in their learning of new speech sounds from a third, unfamiliar language. While it remains to be investigated how experience with two phonological systems might have facilitated learning in the subsample of early bilinguals who were ALs, the behavioral data presented here generally suggest that learning new speech sounds is more significantly influenced by biologically driven neural function than by environment.

The behavioral data also showed that ALs were significantly better at accurately identifying pseudowords as belonging to the same category than their non-AL peers. That is, ALs were behaviorally better than non-ALs at ignoring irrelevant differences between exemplars of the same phoneme. On the other hand, ALs and non-ALs were equally able to notice the phonemic changes that exist between sounds of two different categories. Taken together, these results suggest that ALs are able to tune into novel speech sound contrasts from an unknown language more accurately than their non-AL peers. The ability to learn novel speech sound contrasts likely contributes to relatively more accurate speech and language learning. As Iverson et al. (2003) note, second language learners who are able to identify the *relevant* cues accurately have better speech sound differentiation. Our results suggest that individual ability and not just previous language experience contribute to speech sound

learning, significantly. While initially, all acoustic cues are potentially relevant for distinguishing novel speech sounds from an unknown language, learners who are able to tune into the relevant differences of novel phonemic contrasts have an advantage in phonological acquisition over their peers who have less advanced novel speech sound learning abilities.

Whole-brain within-group comparisons showed that ALs (including monolinguals and bilinguals) and bilinguals (including ALs and non-ALs) had increased activity in the bilateral STG after training. In the direct comparisons between groups, ALs appeared more sensitive to the new speech sound contrasts by showing more activity in the ROI of the STG bilaterally compared to non-ALs. This activity observed in ALs was also more intense and widespread than the brain activity observed for bilinguals relative to monolinguals. Because of the observed neural activity in ALs and bilinguals, two additional contrasts were conducted to tease apart the activity of bilinguals relative to monolinguals in the subset of ALs and separately in the subset of non-ALs. While bilingual ALs and non-ALs showed increased activity in response to novel speech sounds relative to all monolingual learners after training, the results also showed that bilingual ALs had more focused and intense activity in the STG than bilingual non-ALs. As hypothesized, bilingual non-ALs showed two small and weak clusters of activation in the STG bilaterally.

Similarly to the results of performance across language groups, both bilinguals and ALs showed increased activity in the ROIs of the STG in response to within-category and between-category stimuli compared to monolinguals and non-ALs, respectively. In line with the behavioral analysis, ALs showed greater activation in response to between-category stimuli than to within-category stimuli. These results appear to indicate that bilinguals and ALs more readily learn between-category stimuli. Together, these results suggest that while bilingualism may promote learning of novel speech sounds through experience, individual ability contributes significantly to the fundamental underlying platform needed to learn novel speech sounds.

The extraction of cluster values from each ROI per subject further corroborated that ALs were more neurally responsive than non-ALs to the novel speech sounds after training. While bilingual ALs showed greater brain activity than monolingual ALs, this phonological experience in bilinguals appears to have only slightly accentuated the differences between the groups. The main effect is, as shown in the analysis, exerted by individual ability.

The neuroimaging data thus suggest that ALs rely on intense and focalized activation of the STG bilaterally to learn novel speech sounds in general, whereas non-ALs engage this same region to a lesser extent. In the present data, non-ALs did not show any activity above the threshold stipulated in each ROI. The left posterior STG is known to be involved in the early processing of speech (Binder, 2000; Zatorre, Evans, Meyer, & Gjedde, 1992) and has been reported in studies of successful learning after phonetic training (Golestani & Zatorre, 2004; Wong et al., 2007). Additionally, the right STG has been associated with the processing of speech rhythm and prosody (Friederici & Alter, 2004; Geiser, Zaehle, Jancke, & Meyer, 2007). This indicates that ALs are not only good at discriminating the speech sounds embedded in the pseudowords, but also good at noticing the differences in intonation and

overall rhythmic pattern of the pseudowords. ALs, for example, showed much more widespread activity in the right STG than non-ALs. As seen in the whole-brain within-group analysis, ALs show activity in the ventral regions of the caudate nucleus and thalamus, which appear to interconnect the temporal lobes across hemispheres after training. A future functional connectivity analysis or a different method of image acquisition like diffusion tensor imaging could elucidate the notion that ALs have more interconnected superior temporal gyri than non-ALs. Xiang et al. (2012), for example, showed a correlation between language aptitude and temporal inter-hemispheric connectivity.

Some studies have noted that neural activity in response to stimuli to be learned may appear before behavioral improvements are observed (Atienza, Cantero, & Dominguez-Marin, 2002; Tremblay, Kraus, & McGee, 1998). While the results presented here suggest that individual ability exerts a stronger effect than phonological experience in speech sound learning, it is possible that the effect of bilingualism is obscured by the short duration of the training period or by other cognitive and environmental variables not assessed here.

As mentioned previously, the results observed in ALs are consistent with studies that report strong activity of the STG in those who successfully learn after training across different domains, such as music, pitch, phonemes, and words (Bengtsson et al., 2005; Simos et al., 2002; Stewart et al., 2003; Wong et al., 2007). The results observed in ALs suggest that proper learning of speech sounds occurs when sensory processes are highly engaged. Some studies have demonstrated that auditory learning can occur in lower states of consciousness when sensory-based auditory processes replay during sleep (Dave & Margoliash, 2000) or when phonetic training is conducted without feedback and implicit learning mechanisms are primed (Lim & Holt, 2011; Tricomi et al., 2006). The results from these studies, along with the findings presented here, suggest that proper speech learning occurs because sensory processes in the brain are intensely engaged without the interruption of higher-order executive processes, thereby allowing sensory/primary auditory regions of the brain to execute the perceptual task at hand. In other words, learning skills that fundamentally rely on perceptual abilities, like certain aspects of speech acquisition, may be better learned through sensory processing than through higher-order cognitive processing. While the data analysis in the current study cannot posit on other brain areas outside the ROIs selected, the results showed that non-ALs did not engage sensory auditory regions the way ALs did. The recruitment of prefrontal regions, for example, has been observed in subjects experiencing difficulties acquiring speech (Wong et al., 2007) and other skills, such as reading (Temple et al., 2001).

The behavioral and neural data of this study indicate that individual differences in addition to experience contribute significantly to acquiring the phonological system of a novel language. Moreover, the ability to acquire novel speech sound contrasts exists in individuals irrespective of whether they are monolingual or bilingual. Akin to the results reported by Golestani and colleagues (Golestani & Zatorre, 2009; Golestani et al., 2002, 2007, 2011) in the realm of speech and the results of Gaser and Schlaug (2003) in the realm of music, the findings presented here suggest that the neurofunctional contribution to speech learning is significant, and in some respects it may outweigh environmental factors. While bilingualism may not hinder novel speech learning, it does not appear to be advancing such learning

either. A significant effect of bilingualism on speech learning may be observed with longer training or after continuous exposure to the new speech sounds.

In conclusion, it appears that ALs and non-ALs are normally distributed in the population, regardless of their phonetic experiences. Some individuals appear to have a greater inclination for speech sound learning because their STG is more highly engaged than others'. With the expected growth of bilingual children in the upcoming years worldwide, it is important to understand the interaction between early bilingual environments, genetic predispositions, and neural function as basic speech perception processes can reveal how the interplay of these factors affect phonological development in particular, and language development in general.

Acknowledgments

We want to express our heartfelt gratitude to Mária Gósy of the Hungarian Academy of Sciences, Research Institute for Linguistics and Judit Bóna of the Institute for Hungarian and Linguistics and Finno-Ugric Studies for assisting us in obtaining the Hungarian recordings. We would also like to thank Tom Zeffiro from the Martinos Center for Biomedical Imaging at the Massachusetts General Hospital for his guidance with fMRI analysis, the team of research assistants who helped collect the data for this project, and Kailyn Bradley for comments on previous versions of this manuscript.

Funding: This work was supported by the National Institute of Health (NIH) for the Neural correlates of lexical processing in child L2 learners (1R21HD059103-01A1).

References

- Archila-Suerte P, Zevin J, Ramos A, Hernandez A. The neural basis of non-native speech perception in bilingual children. *NeuroImage*. 2013; 67:51–63. [PubMed: 23123633]
- Atienza M, Cantero JL, Dominguez-Marin E. The time course of neural changes underlying auditory perceptual learning. *Learning & Memory*. 2002; 9(3):138–150. [PubMed: 12075002]
- Beach E, Burnham D, Kitamura C. Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *International Journal of Bilingualism*. 2001; 5(2):221–235.
- Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H, Ullen F. Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*. 2005; 8(9):1148–1150. [PubMed: 16116456]
- Best CC, McRoberts GW. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*. 2003; 46(Pt 2–3):183–216. [PubMed: 14748444]
- Binder J. The new neuroanatomy of speech perception. *Brain*. 2000; 123(Pt 12):2371–2372. [PubMed: 11099441]
- Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L. Deviant processing of letters and speech sounds as proximate cause of reading failure: A functional magnetic resonance imaging study of dyslexic children. *Brain*. 2010; 133(Pt 3):868–879. [PubMed: 20061325]
- Burgaleta M, Baus C, Diaz B, Sebastian-Galles N. Brain structure is related to speech perception abilities in bilinguals. *Brain Structure & Function*. 2014; 219(4):1405–1416. [PubMed: 23686398]
- Callan D, Tajima K, Callan A, Akahane-Yamada R, Masaki S. Neural processes underlying perceptual learning of a difficult second language phonetic contrast. Paper presented at the Proceedings of Eurospeech; Aalborg. 2001.
- Callan D, Tajima K, Callan A, Kubo R, Masaki S, Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*. 2003; 19(1):113–124. [PubMed: 12781731]

- Chandrasekaran B, Sampath PD, Wong PC. Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*. 2010; 128(1):456–465. [PubMed: 20649239]
- Cohen J, Flatt M, MacWhinney B, Provost J. *PsyScope X Build 57*. Pittsburgh: 2010.
- Cohen SP, Tucker GR, Lambert WE. The comparative skills of monolinguals and bilinguals in perceiving phoneme sequences. *Language and Speech*. 1967; 10(3):159–168. [PubMed: 5583882]
- Dave AS, Margoliash D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*. 2000; 290(5492):812–816. [PubMed: 11052946]
- Diaz B, Baus C, Escera C, Costa A, Sebastian-Galles N. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(42):16083–16088. [PubMed: 18852470]
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*. 2005; 25(4):1325–1335. [PubMed: 15850749]
- Enomoto K. *Edinburgh Working Papers in Applied Linguistics*. Vol. 5. Edinburgh, UK: Edinburgh Working Papers in Applied Linguistics; 1994. L2 perceptual acquisition: The effect of multilingual linguistic experience on the perception of a "less novel" contrast; 15–29.
- Flege JE. Assessing constraints on second language segmental production and perception. In: Schiller NO, Meyer AS, editors *Phonetics and phonology in language comprehension and production*. Berlin, Germany: Mouton de Gruyter; 2003. 319–355.
- Friederici AD, Alter K. Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*. 2004; 89(2):267–276. [PubMed: 15068909]
- Gaab N, Gaser C, Schlaug G. Improvement-related functional plasticity following pitch memory training. *NeuroImage*. 2006; 31(1):255–263. [PubMed: 16427320]
- Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*. 2003; 23(27):9240–9245. [PubMed: 14534258]
- Geiser E, Zaehle T, Jancke L, Meyer M. The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*. 2007; 20(3):541–552.
- Golestani N, Molko N, Dehaene S, LeBihan D, Pallier C. Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*. 2007; 17(3):575–582. [PubMed: 16603709]
- Golestani N, Paus T, Zatorre RJ. Anatomical correlates of learning novel speech sounds. *Neuron*. 2002; 35(5):997–1010. [PubMed: 12372292]
- Golestani N, Price CJ, Scott SK. Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience*. 2011; 31(11):4213–4220. [PubMed: 21411662]
- Golestani N, Zatorre RJ. Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*. 2004; 21(2):494–506. [PubMed: 14980552]
- Golestani N, Zatorre RJ. Individual differences in the acquisition of second language phonology. *Brain and Language*. 2009; 109(2–3):55–67. [PubMed: 18295875]
- Iverson P, Hazan V, Bannister K. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*. 2005; 118(5):3267–3278. [PubMed: 16334698]
- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, Siebert C. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 2003; 87(1):B47–57. [PubMed: 12499111]
- Joanisse MF, Zevin JD, McCandliss BD. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*. 2007; 17(9):2084–2093. [PubMed: 17138597]
- Lim SJ, Holt LL. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*. 2011; 35(7):1390–1405. [PubMed: 21827533]
- Maldjian JA, Laurienti PJ, Burdette JH. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *NeuroImage*. 2004; 21(1):450–455. [PubMed: 14741682]

- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*. 2003; 19(3):1233–1239. [PubMed: 12880848]
- McCandliss B, Fiez J, Protopapas A, Conway M, McClelland J. Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*. 2002; 2(2):89–108.
- Nittrouer S, Crowther CS, Miller ME. The relative weighting of acoustic properties in the perception of [s] + stop clusters by children and adults. *Perception & Psychophysics*. 1998; 60(1):51–64. [PubMed: 9503911]
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971; 9(1):97–113. [PubMed: 5146491]
- Pallier C, Bosch L, Sebastian-Galles N. A limit on behavioral plasticity in speech perception. *Cognition*. 1997; 64(3):B9–17. [PubMed: 9426508]
- Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. Increased auditory cortical representation in musicians. *Nature*. 1998; 392(6678):811–814. [PubMed: 9572139]
- Pisoni D, Tash J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*. 1974; 15(2):285–290. [PubMed: 23226881]
- Pruitt J, Jenkins JJ, Strange W. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America*. 2005; 119(3):1684–1696.
- Ressel V, Pallier C, Ventura-Campos N, Diaz B, Roessler A, Avila C, Sebastian-Galles N. An effect of bilingualism on the auditory cortex. *Journal of Neuroscience*. 2012; 32(47):16597–16601. [PubMed: 23175815]
- Sebastián-Gallés N, Soriano-Mas C, Baus C, Díaz B, Ressel V, Pallier C, ... Pujol J. Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics*. 2012; 25(3):150–162.
- Simos PG, Fletcher JM, Bergman E, Breier JJ, Foorman BR, Castillo EM, ... Papanicolaou AC. Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology*. 2002; 58(8):1203–1213. [PubMed: 11971088]
- Stewart L, Henson R, Kampe K, Walsh V, Turner R, Frith U. Brain changes after learning to read and play music. *NeuroImage*. 2003; 20(1):71–83. [PubMed: 14527571]
- Temple E, Poldrack RA, Salidis J, Deutsch GK, Tallal P, Merzenich MM, Gabrieli JDE. Disrupted neural responses to phonological and orthographic processing in dyslexic children: An fMRI study. *NeuroReport*. 2001; 12(2):299–307. [PubMed: 11209939]
- Tremblay K, Kraus N, McGee T. The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport*. 1998; 9(16):3557–3560. [PubMed: 9858359]
- Tricoli E, Delgado MR, McCandliss BD, McClelland JL, Fiez JA. Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*. 2006; 18(6):1029–1043. [PubMed: 16839308]
- Werker JF. The effect of multilingualism on phonetic perceptual experience. *Applied Psycholinguistics*. 1986; 7:141–156.
- Wong PCM, Perrachione TK, Parrish TB. Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*. 2007; 28(10):995–1006. [PubMed: 17133399]
- Wood CC. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *The Journal of the Acoustical Society of America*. 1976; 60(6):1381–1389. [PubMed: 1010890]
- Woodcock RW. Woodcock language proficiency battery – Revised (WLPB-R). Camberwell, Australia: Riverside Publishing; 1995.
- Woodcock RW, Muñoz-Sandoval A. Woodcock-Johnson language proficiency battery – Revised (Spanish). Itasca, IL: The Riverside Company; 1995.
- Xiang H, Dediu D, Roberts L, Van Oort E, Norris DG, Hagoort P. The structural connectivity underpinning language aptitude, working memory, and IQ in the Perisylvian language network. *Language Learning*. 2012; 62:110–130.

- Zatorre R, Evans A, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science*. 1992; 256(5058):846–849. [PubMed: 1589767]
- Zhang Y, Kuhl P, Imada T, Iverson P, Pruitt J, Kotani M, Stevens E. Neural plasticity revealed in perceptual training of a Japanese adult listener to learn American /l-r/ contrast: A whole-head magnetoencephalography study. Paper presented at the Sixth International Conference on Spoken Language Processing; 2000.
- Zhang Y, Kuhl PK, Imada T, Iverson P, Pruitt J, Stevens EB, ... Nemoto I. Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*. 2009; 46(1):226–240. [PubMed: 19457395]

Biographies

Pilar Archila-Suerte is a postdoctoral fellow in Psychology at the University of Houston. Her research interests include the development of speech perception in bilinguals, bilingual language development, and the effects of bilingualism on general cognitive processing.

Ferenc Bunta is an associate professor in Communication Sciences and Disorders at the University of Houston in Houston, Texas. His research focuses on cross-linguistic and bilingual phonological acquisition in children and adults with typical speech, language, and hearing as well as children with hearing loss.

Arturo E. Hernandez is a professor and director of Developmental Psychology at the University of Houston and director of the Laboratory for the Neural Bases of Bilingualism. His research agenda focuses on the nature of bilingual language processing in word learning and cognitive control using both behavioral and neuroimaging methods.

Figure 1. Vowel chart of Spanish, English and target Hungarian sounds

Adapted from the International Phonetic Alphabet (IPA). Spanish vowel sound symbols are color-coded in red, English in green, and Hungarian in blue. Note that Hungarian vowels are unique in their distribution with respect to English and Spanish vowels. The symbols for Hungarian sounds are represented by the following letters: y = ü; y: = ; ø = ö; ø: = ; ɔ: = ó; ɒ = o.

Figure 2.

(A) Bar graphs of mean language proficiency. Mean proficiency of English in monolinguals and bilinguals and mean proficiency of English and Spanish in bilinguals. Error bars indicate standard error. (B) Progression of learning in monolinguals and bilinguals. Monolinguals and bilinguals significantly improved during the training but the groups did not differ from each other learning novel speech sounds.

Figure 3(a)

Figure 3(b)

Figure 3.

Figure 3(a). Within-subject variability across posttests.

Only half the participants are presented in this graph in order to have a more understandable and cleaner illustration. The encircled peaks highlight some of the participants that performed more accurately during the 3rd posttest than in the last posttest. To emphasize the scores of the last two posttests, the line representing the 3rd posttest is marked in bold and dotted and the line representing the 4th posttest is marked in bold and straight. The lines representing post-tests 1 and 2 are dotted and dashed lighter gray.

Figure 3(b). Performance groups given hierarchical clustering classification.

Note: This illustration shows the distribution of advanced and non-advanced learners given their mean score performance across posttests (represented in the y -axis) and English proficiency score (represented in the x -axis).

Figure 4(a)

Figure 4(b)

Figure 4.

Figure 4(a). Neural activity in bilateral STG ROI in response to novel speech sounds vs. baseline.

Bilinguals show increased activity relative to monolinguals in response to novel speech versus baseline after training. Similarly, ALs show increased activity relative to non-ALs in response to novel speech versus baseline after training. See Table 1 for a list of intensity values, cluster sizes, and MNI coordinates.

STG: superior temporal gyrus; ROI: region of interest; AL: advanced learner

Figure 4(b). Neural activity in bilateral STG ROI in response to between-category stimuli vs. baseline.

Bilinguals show increased activity relative to monolinguals in response to between-category stimuli versus baseline after training. Similarly, ALs show increased activity relative to non-ALs in response to between-category stimuli versus baseline after training. See Table 1 for a list of intensity values, cluster sizes, and MNI coordinates.

STG: superior temporal gyrus; ROI: region of interest; AL: advanced learner

Figure 5(a)

Figure 5(b)

Figure 5.

Figure 5(a). Performance main effect in left STG ROI for cluster magnitude.

Advanced learners showed larger clusters of activation than non-advanced learners in the ROI of the left STG. STG: superior temporal gyrus; ROI: region of interest

Figure 5(b). Performance main effect in right STG ROI for cluster magnitude.

Advanced learners showed larger clusters of activation than non-advanced learners in the ROI of the right

STG. STG: superior temporal gyrus; ROI: region of interest

Table 1

Whole-brain analysis				
Hemisphere	Brain area	Cluster size (#voxels)	Peak <i>T</i>	MNI coordinates
<i>Pre-training fMRI in monolinguals</i>				
Left	STG	497	6.19	−42 −22 2
Right	STG	552	5.79	48 −12 −2
<i>Pre-training fMRI in bilinguals</i>				
Right	STG	21	4.76	60 −20 −2
<i>Post-training fMRI in bilinguals</i>				
Right	STG	1084	6.01	54 −10 −2
Left	STG	679	5.7	−50 −16 4
Left	STG/Insula	21	4.94	−38 −42 22
Right	Cingulate gyrus/ precuneus	21	4.79	22 −40 28
<i>Pre-training fMRI in advanced learners</i>				
Left	STG	995	6.41	−44 −18 6
Right	STG	1006	5.9	50 −10 −2
<i>Post-training fMRI in non-advanced learners</i>				
Left	STG	1232	6.62	−44 −22 −2
Right	STG	1152	6.44	54 −6 −2
Right	Caudate/thalamus	261	5.62	12 −36 14
Right	Caudate/thalamus	158	5.4	4 −10 20
Left	Posterior cingulate	30	4.93	−12 −42 16
Region of interest analysis – superior temporal gyrus				
Hemisphere	Cluster size (#voxels)	Peak <i>T</i>	MNI coordinates	
<i>Post-training advanced learners >non-advanced learners</i>				
Left	290	2.84	−42 −20 −4	
Left	201	2.75	−60 −44 14	
<i>Post-training advanced learners >non-advanced learners</i>				
Right	113	2.55	44 0 −12	
Right	178	2.28	68 −36 22	
Right	31	2.23	68 −32 4	
Contrasts within performance and across language				
<i>Post-training advanced bilingual learners >advanced monolingual learners</i>				
Right	331	2.45	42 −38 4	
<i>Post-training non-advanced bilingual learners >non-advanced monolingual learners</i>				
Left	39	2.05	−40 −38 16	
Left	28	2.03	−64 −32 10	
Right	68	2.23	66 −36 10	
Right	22	2.11	42 −28 6	
Within-category and between-category comparisons				

Whole-brain analysis				
Hemisphere	Brain area	Cluster size (#voxels)	Peak <i>T</i>	MNI coordinates
<i>Bilingual > monolingual</i>				
<i>Within-category</i>				
Left	51	2.19	−48 −20 12	
Left	45	2.05	−48 −32 14	
Right	247	2.19	56 −14 −2	
<i>Between-category</i>				
Left	786	2.58	−50 −18 12	
Right	609	2.53	44 −26 0	
<i>Advanced learner > Non-advanced learner</i>				
<i>Within-category</i>				
Left	127	2.44	−62 −42 16	
Left	23	2.07	−40 −20 −2	
Left	44	2.05	−48 0 −2	
<i>Between-category</i>				
Left	444	3.36	−40 −16 −6	
Left	267	2.87	−60 −44 14	
Left	29	2.15	−38 −38 18	
Right	302	3.14	44 0 −12	
Right	523	2.54	68 −34 6	

STG: superior temporal gyrus