Language Experience in Second Language Speech Learning

In honor of James Emil Flege

Edited by

Ocke-Schwen Bohn
Aarhus University

Murray J. Munro
Simon Fraser University

John Benjamins Publishing Company
Amsterdam / Philadelphia

2007

CHAPTER 14

Behavioral and cortical effects of learning a second language

The acquisition of tone

Joan A. Sereno and Yue Wang

Explanations of language learning often involve appeals to distinct learning mechanisms. On one hand, learners' innate characteristics are emphasized, with learning tied to a limited time period, a critical period, when the brain is predisposed for success in language learning. This view is often contrasted with a position emphasizing the role of the environment in shaping language learning, highlighting the contribution of feedback mechanisms and the nature of the speech input. One approach has been to examine how second languages are learned in order to directly examine change due to learning. To this end, the present paper documents the behavioral and cortical changes resulting from learning a novel language contrast, specifically Mandarin tone. Hemispheric differences in the processing of language contrasts are observed, with significant left hemispheric participation in native listeners and no hemispheric preference for non-native listeners. Additional experiments examined the training of non-native listeners, revealing that tone perception accuracy can be improved with minimal exposure. Furthermore, it can be generalized to new stimuli and talkers, retained for at least six months, and transferred to production. Native listeners identify post-training productions more accurately than pre-training productions and acoustic analyses of the post-training F0 contours show better approximation to native speaker norms. These behavioral changes due to training can also be observed cortically, with the learning of Mandarin tone contrasts associated with significant increases in activity in the traditional language areas (left hemisphere superior temporal gyrus) as well as the recruitment of neighboring neural areas. Implications for theories of language learning will be addressed.

Introduction

Explanations of second language learning often involve an appeal to distinct learning mechanisms. Many of these proposed explanations have been molded by research in first language acquisition. In first language acquisition, theories of learning often contrast two positions. On the one hand, learners' innate characteristics are emphasized, with learning tied to a specific and limited time period, a critical period, when the brain is predisposed for success in language learning. This view is often contrasted with a position emphasizing the role of the environment in shaping language learning, highlighting the contribution of feedback mechanisms and child-directed speech input. Parallel arguments are evident in second language learning. The present paper will contribute to this discussion by examining aspects of how languages are learned.

Second languages are learned, albeit often imperfectly. Regardless of one's theoretical emphases, one observation consistently holds: That is, those individuals who begin learning a second language late in adolescence (late bilinguals) differ from those who begin learning a second language earlier in adolescence (early bilinguals). While children of immigrant families often speak a second language with native-like fluency, their parents rarely achieve such levels of mastery. Differences between these participant groups cross a number of linguistic domains, from marked variation in accent and word choice to deviations in morphology or syntax. Explanations for these divergences range from biological to social descriptions. Biological explanations claim that late bilinguals' brains are less plastic and therefore learning is easier in younger learners. Linguistic explanations appeal to the fact that the late bilinguals' well-established native language system influences the developing second language (L2) system, resulting in the differential acquisition of similar and new phonemes. Late bilinguals may also receive less or, at least, different native speaker input. In such environmental explanations, younger learners have more time to devote to language learning and more opportunities to hear and use the language. Finally, some explanations appeal to differences in motivation. Younger learners regularly use language in an informal language learning environment in which there is less pressure to speak fluently and errors are often tolerated. Older learners' language situation demands much more complex language which is often accompanied by frustration and the embarrassment of not being able to communicate adequately. Needless to say, no definite evidence is available to distinguish among these competing theories yet. One step in the right direction is to examine how first languages are processed and how second languages are learned in order to directly examine changes as a result of language learning.

Mandarin lexical tone

Our approach to these issues has been to investigate in detail the processing of lexical tone (for a review, see Wang, 2001, and Wang, Jongman, & Sereno, 2006). Listeners of tone languages (e.g., Mandarin Chinese) use a combination of fundamental frequency and durational cues (for a review, see Jongman, Wang, Moore, & Sereno, 2005). Mandarin Chinese phonemically distinguishes four tones, with Tone 1 having high-level pitch, Tone 2 high-rising pitch, Tone 3 low-dipping pitch, and Tone 4 high-falling pitch. Figure 1 shows fundamental frequency contours (F0 and duration information) for the four Mandarin tones, each spoken as the syllable *ma* produced by a native Mandarin speaker.

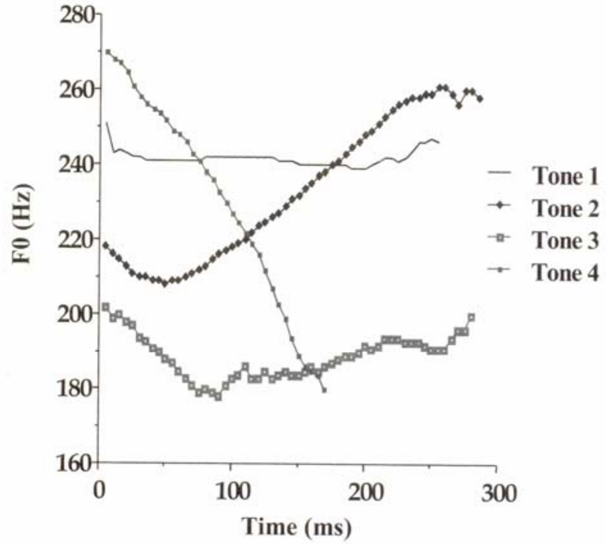


Figure 1. F0 contours (Hz) for each of the four Mandarin Chinese tones for the segmental context *ma* spoken in isolation by a female speaker (from Moore & Jongman, 1997)

Tone processing has important implications at the cortical level. Previous research indicates that the left hemisphere is more adept at phonemic processing, including phonemes, syllables and words (Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970) while the right hemisphere is better at melodic and prosodic processing, including music, pitch contours, and affective prosody (Kimura, 1964; Curry, 1967; Bryden, 1963). While tones are used to make phonemic contrasts in tone languages (often a left hemisphere function), tones can also be characterized as involving a modulation of F0 (generally assumed to be the domain of the right hemisphere). Consequently, lexical tone is a useful medium for studying hemispheric specialization.

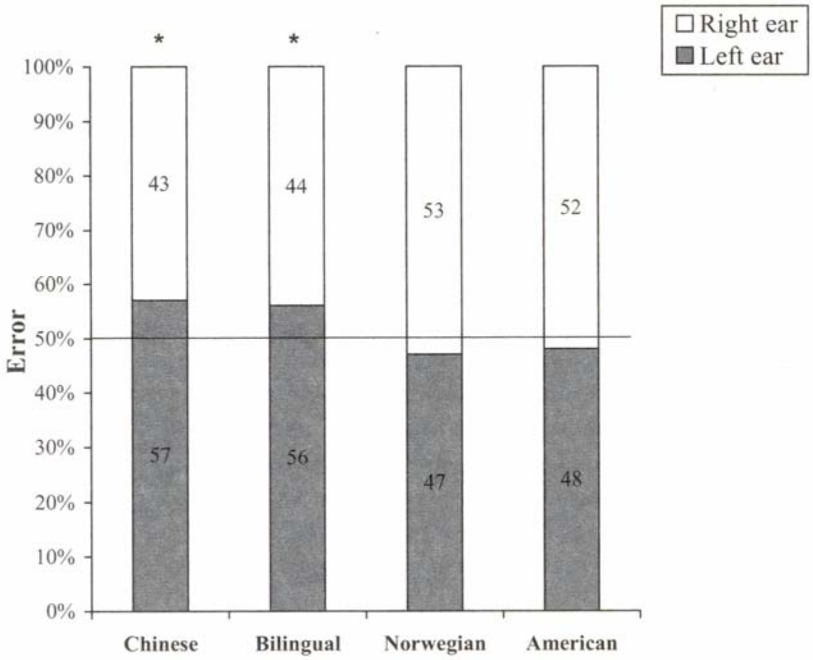


Figure 2. Distribution of left-ear errors and right-ear errors (in %) for the Chinese (n=20), bilingual (n=15), Norwegian (n=20), and American (n=20) listeners. *indicates significance at $p<.05$ (From Wang, Behne, Jongman, & Sereno, 1999)

Hemispheric processing and tone: native listeners

Dichotic listening provides a research paradigm to investigate possible functional lateralization of tone. Since the right ear is primarily connected to the left hemisphere and the left ear to the right hemisphere via contralateral pathways, errors across the ears, when there is simultaneous dichotic presentation, can provide information on how listeners process stimuli. Research using dichotic presentation has consistently shown a right ear advantage (i.e., a left hemispheric dominance) for linguistic stimuli. A right ear advantage (REA) has generally been found as well for tone processing (e.g., Van Lancker & Fromkin, 1973). These data have been reported for both Thai and Norwegian listeners but inconsistent data have been reported for Mandarin Chinese. In a dichotic listening study, we (Wang, Jongman, & Sereno, 2001) investigated lateralization of lexical tone in native Mandarin listeners. Twenty Chinese listeners were asked to identify dichotically presented tone pairs by identifying which tone they heard in each ear. To induce errors, stimuli were embedded in white noise (-10 db signal-to-noise, S/N, ratio) with a 2-second inter-stimulus-interval. The stimuli were 16 commonly used monosyllabic Mandarin words, consisting of 4 different stimuli each combined with the four tones, resulting in 4 quadruplets. The results showed more errors in the left ear than in the right ear (an REA), demonstrating a significant left hemispheric advantage of the processing of Mandarin tones by native Mandarin listeners. Figure 2 (leftmost bar, "Chinese") displays listeners' performance in terms of percentage of left (57%) and right (43%) ear errors.

This REA for processing tone occurred in most listeners, with 15 out of 20 Mandarin listeners exhibiting an REA. Data were also analyzed in terms of individual tones. While the overall number of errors for tone 3 was significantly larger than for the other three tones, a sizeable REA was observed across all 4 tones. For native listeners, the perception of tonal contrasts in a tone language is to a large extent a property of the left hemisphere.

Hemispheric processing and tone: non-native listeners

To further investigate the hemispheric processing of tone, additional listener groups were examined. Listener groups included American English listeners without any tone language background, Norwegian listeners with tone experience in their native language but no experience in Mandarin, and Chinese-English bilinguals who are fluent in both Mandarin and English. Each of these listener groups will be discussed in turn.

We (Wang et al., 2001) initially investigated lateralization of lexical tone in American listeners to examine how listeners of languages which do not make tonal distinctions process tone. These American listeners had no knowledge of Mandarin Chinese or any other tonal language. As a result, the American listeners received brief instructions (about 30 minutes) in order to familiarize them with the 4 Mandarin tones prior to testing. Similar to the procedures used with the native Mandarin speakers, twenty American listeners were asked to identify dichotically presented tone pairs by identifying which tone they heard in each ear. To insure a comparable number of overall errors to the native Mandarin listeners, stimuli were embedded in white noise (0 db S/N ratio) with a longer inter-stimulus-interval (4 seconds). All other testing procedures were identical to the testing of the native speakers. For these non-native listeners (American listeners), a quite different pattern emerges. While native speakers of a tone language show an REA when processing tone stimuli, these non-native listeners (American listeners) show no clear ear asymmetry (see Figure 2, rightmost bar "American"). The American listeners showed no significant differences in the processing of tonal stimuli across the hemispheres as indicated by a comparable number of errors in the left (48%) and right (52%) ears. Moreover, this lack of a significant REA in the processing of Mandarin tones by non-natives was consistent across listeners, with only 8 out of 20 American listeners showing an effect. Data were also analyzed in terms of individual tones and no ear advantage was observed for any individual tone. However, there were some differences observed in number of overall errors across specific tones and this pattern did not mimic that found for the native listeners. For the non-native listeners, tone 4 was the most difficult and tones 1 and 3 were the easiest, while errors for native listeners were most frequent for tone 3. Overall, then, given the consistent REA for native listeners and lack of one for the American English non-native listeners, the perception of tonal contrasts in a tone language seems to a large extent to be a property of the left hemisphere. For non-native listeners, tonal contrasts do not carry linguistically significant information and consequently no left hemispheric superiority is observed.

A second group of non-native listeners were examined. We (Wang, Behne, Jongman, & Sereno, 2004) investigated lateralization of lexical tone in Norwegian listeners. Norwegian presents an interesting extension in that Norwegian has tonal contrasts. Some dialects of Norwegian maintain a tonal distinction, sometimes referred to as tonemes (see for example, Kristoffersen, 2000). In East Norwegian, for example, toneme 1 is a low tone while toneme 2 consists of a high tone falling to a low tone. Norwegian therefore has a tonal contrast, but it manifests itself differently than in Mandarin. Nevertheless, similar to Mandarin tone processing, previous research found that the Norwegian tones are primarily processed in the left hemisphere by native listeners of Norwegian (Moen, 1993). The question arises, then, whether Norwegian native speakers would process Mandarin tone primarily in the left hemisphere since tone has linguistic significance in their native language or whether these listeners would process tone as generic pitch distinctions and show no lateralization as listeners of non-tone languages do. To investigate this question, we (Wang et al., 2004) tested Norwegian listeners who had no knowledge of Mandarin Chinese. Twenty Norwegian listeners were asked to identify dichotically presented tone pairs by identifying which tone they heard in each ear. Identical procedures used with the non-native speakers (American listeners) were employed to insure a comparable number of overall errors (0 db S/N ratio in white noise; 4 seconds ISI). All other testing procedures were identical to the earlier testing of the native speakers. We found that while native listeners of a tone language show an REA when processing tone stimuli, the Norwegian non-native listeners show no clear ear asymmetry, similar to the non-tonal language American listeners. The Norwegian listeners showed no differences in processing of tonal stimuli across the hemispheres, exhibiting a comparable number of errors in the left (47%) and right (53%) ears (see Figure 2, "Norwegian"). In fact, only 7 out the 20 listeners showed an REA. Individual tone analyses also showed no ear advantage. Similar to the pattern observed for the non-native, non-tonal language users, the overall number of errors across specific tones showed a greater number of errors for tone 4 as compared to tones 1, 2, and 3. Again, the lack of a significant REA and the pattern of errors across the four tones were parallel to that observed for the non-native American listeners. It seems that experience with a language that makes tonal distinctions did not affect the hemispheric processing of tonal distinctions in an unfamiliar language. Even though Norwegian tones are primarily processed in the left hemisphere by native speakers of Norwegian, when processing unfamiliar Mandarin Chinese, Norwegian listeners do not show left hemispheric superiority. Instead, the Norwegian listeners are similar to the American listeners, with neither group showing any preference in hemispheric processing of Mandarin tones. These results suggest a language-dependent lateralization effect, with left hemispheric specialization only present when the tonal distinctions are linguistically significant to the listener.

To complete the picture, a final listener group was examined. We (Wang et al., 2004) investigated a group of highly-proficient bilingual listeners in order to evaluate whether the hemispheric specialization of bilingual listeners who were exposed to a tone language as a second language (L2) differs from that of native listeners. Lateralization of language processing by bilinguals and L2 learners has been quite controversial in the literature. While some studies show that the languages of bilinguals are left-lateralized similar to monolinguals (e.g., Soares, 1982, 1984; Soares & Grosjean, 1981), others have shown greater right hemispheric involvement in language processing in bilinguals (Mildner, 1999) or even equal processing across both hemispheres in bilinguals (Ke, 1992). When age of L2 acquisition is manipulated, in general, studies have demonstrated left-hemispheric dominance for early bilinguals and symmetrical hemispheric involvement for later bilinguals (Sussman, Franklin, & Simon, 1982; Wuillemin & Richardson, 1994). However, manipulation of proficiency shows slightly different results, with growing L2 proficiency generally showing increased right hemispheric participation (Albanese, 1985; but see Schouten, van Dalen, & Klein 1985).

For these reasons, the investigation of lateralization of tone in bilinguals is of interest. Our English-Mandarin bilingual listeners were born and raised in the United States. All claimed to have acquired English as their first and dominant language. Since they all had native Chinese parents or family members, they had been exposed to Mandarin (speaking and/or comprehension) from birth. They were fluent in Mandarin as well as English. Identical testing procedures were used for the native and bilingual listeners. The results show that the English-Mandarin bilingual listeners exhibit a left-hemispheric superiority (see Figure 2, "Bilingual"). The data for the English-Mandarin bilinguals follow the same pattern shown by the native Mandarin listeners, a significant REA for processing the tonal stimuli, with significantly more left ear (56%) than right ear (44%) errors. This effect held across participants, with 12 out of 15 bilinguals making more errors on the left ear than on the right ear. Consistent with the overall result showing an REA for each of the individual tones, the mean number of left ear errors is greater than that of the right ear. Moreover, similar to the native listener data across tones, the overall number of errors for tone 3 was significantly greater than the other three tones (tones 1, 2 and 4). It seems that the proficient bilinguals do not differ from the native Chinese listeners in terms of left hemispheric dominance for processing Mandarin tones. When highly proficient Mandarin-English bilinguals show near native fluency, their tone processing also is native-like. These results for native Mandarin Chinese listeners and bilingual Mandarin listeners are consistent, suggesting a left hemispheric processing of lexical tone.

Together, the dichotic experiments show systematic differences across listener groups, with post hoc analyses showing that the native Mandarin and bilingual listeners were significantly different from the American and Norwegian listeners. Native Mandarin listeners and bilingual English-Mandarin listeners show a significant REA in the perception of Mandarin tone, demonstrating a left hemispheric superiority in the processing of lexical tone by native tone-language speakers. In contrast, non-native listeners show a different pattern. For both American listeners and for Norwegian listeners, who are familiar with Norwegian tones but not Mandarin tones, no ear preference was observed indicating no hemispheric dominance. Neither non-native group processed tone primarily in the left hemisphere, as the native speakers did. Only listeners with extensive prior experience with Mandarin tone exploit left lateralized processes and show right ear preferences in processing Mandarin tone. These data raise the interesting issue whether non-native listeners' tone processing patterns could mimic native-like patterns as the inexperienced listeners gain more experience with Mandarin tones. An important extension of these results, then, is to determine whether speakers of a non-tone language can be trained to process the signal in a manner similar to a tone-language user. The specific question is whether American listeners, for example, can be trained to identify the four Mandarin tones and, behaviorally, approach native-listener accuracy.

Training with tone

Classic early training studies examined the acquisition of novel phonetic categories by non-native speakers, a persistent problem for second language learners trying to acquire a new language contrast. Early reports suggested that even though discrimination of stimuli improved slightly during training, the effects did not generalize to new stimuli (Strange & Dittman, 1984). However, modification of the training procedures using a high variability paradigm (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993) proved to effectively facilitate learning of nonnative speech contrasts. Unlike the earlier attempts, training was accomplished using stimuli from a wide range of phonetic environments and a wide variety of talkers so that listeners formed context-sensitive representations. Using these procedures, it was found that non-native segmental contrasts could be learned and that the adult perceptual system had the capacity to change. Contrasts included a three-way voice onset time distinction (Pisoni, Aslin, Perey, & Hennessy, 1982), the interdental voiced-voiceless distinction (Jamieson & Morosan, 1986, 1989), and the more extensively examined /r/-/l/ distinction (Logan et al., 1991; Lively et al., 1993; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). These studies clearly showed that identification of non-native speech contrasts improved with training, the improvement was extended to novel stimuli from different talkers, and the phonetic contrasts were retained long after training. These training studies demonstrated long-term modification of listeners' segmental phonetic categories.

We (Wang, Spence, Jongman, & Sereno, 1999) extended these training results to the suprasegmental domain, examining learning of Mandarin tonal contrasts by non-native listeners. Generally, for adult non-tonal speakers learning Mandarin as an L2, tone has been a source of difficulty in learning (e.g., Broselow, Hurtig, & Ringen, 1987; Shen, 1989). Since tonal patterns are an integral part of learning words in Mandarin, language learning is incomplete without their mastery.

To investigate tone acquisition, we applied the above-mentioned high variability training procedure to the acquisition of Mandarin tonal contrasts. Sixteen native speakers of English participated, with eight as trainees and eight as controls. Both trainees and controls took a pretest in which they were presented with 100 randomized Mandarin stimuli. Subjects were to identify the tone of each stimulus. Immediately after pretest, the eight trainees participated in a two-week training program consisting of eight sessions of 40 minutes each. The training stimuli appeared in a variety of phonetic contexts in natural words produced by four different talkers. The trainees were to identify the tone for each stimulus they heard. During training, immediate feedback was given after each stimulus. The control subjects did not participate in any of the training sessions. A posttest, identical to the pretest, was then given to both trainee and control subjects.

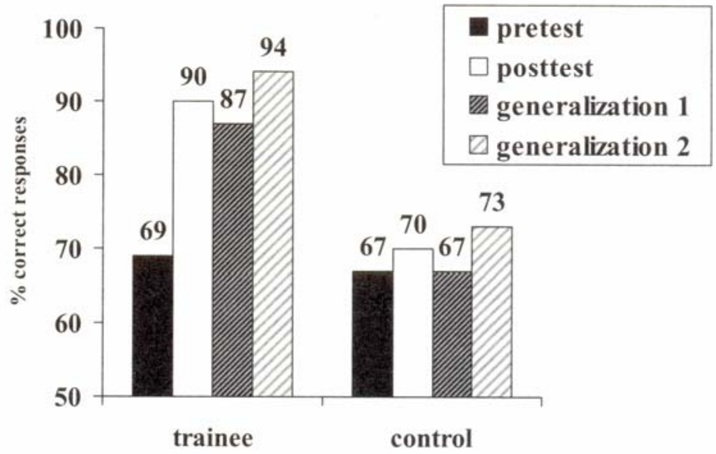


Figure 3. Mean %-correct ID of the tones for trained (n=8) and control (n=8) subjects at pretest, post-test, generalization test 1 (old talker, new stimuli), and generalization test 2 (new talker, new stimuli) (Wang, Spence, Jongman, & Sereno, 1999)

The data (Wang et al., 1999) are presented in Figure 3, showing mean percent correct identification of the four Mandarin tones for the trainee and control subjects. Correct identification scores revealed a substantial improvement in tone identification for the trainees from pretest (69%) to posttest (90%), a dramatic 21% increase in tone identification accuracy. This contrasts with the controls who show no change from pretest (67%) to posttest (70%) performance, a non-significant 3% difference.

Trainees' performance also showed significant percentage point improvement from pretest to posttest for each individual tone, with tone 1 showing a 15% improvement, tone 2 a 22% improvement, tone 3 an 18% improvement, and tone 4 a 25% improvement. In general, the pattern of tone confusions before and after training is quite comparable, with all participants showing similar behavior. Although there is a large degree of variability among the eight trainees' initial pretest accuracy levels, improvement was systematically observed for each individual trainee, ranging from 6% to 33%.

Two additional tests were conducted to evaluate the generalizability of the results to novel stimuli produced by one of the training speakers (Generalization Test 1) and to novel stimuli produced by an entirely new speaker (Generalization Test 2). For both generalization tests, tone identification scores differed across trainees and controls. As shown in Figure 3, the controls showed no difference between pretest, posttest, generalization test 1, or generalization test 2. Without training, controls showed no improvement. In contrast, the trainees showed significant improvement in accuracy relative to pretest scores, both to new stimuli (an 18% increase) and to new talkers and stimuli (a 25% increase), as shown in Figure 3. A final evaluation of the training paradigm was to determine whether the increase in identification accuracy of the tonal contrasts would persist over time. A subset of eight subjects participated six months after training. Four trainees still showed significant improvement in tone identification (87% correct identification) while four control participants showed consistently low accuracy (58% correct identification).

In sum, these training data suggest that perception of Mandarin tones can be improved using a simple training task, showing a robust 21% increase in trainees' overall tone perception accuracy. This held for all four tones and it was observed for all individual trainees. In addition, the improvement gained in training generalized to new stimuli and new talkers and was retained by listeners six months after training. These results support the notion that this high variability training procedure facilitated the formation of new phonetic categories. For English speakers, the association between segmental structure and F0 contour, as exhibited in linguistic tone patterns, does not exist. But given appropriate training, these second language learners can learn to perceive the tonal distinctions. The adult perceptual system can be modified.

Cortical modification during tone learning

While our data show that learners become more native-like behaviorally, the question remains whether we might be able to observe these behavioral improvements at a cortical level. Our dichotic listening data discussed above suggest that the neural substrates underlying the ability to identify lexical tone are predominantly lateralized in the left hemisphere. While native speakers of tone languages show left hemispheric specialization for the processing of tone, non-native speakers do not. The training procedures used with non-native tone learners, however, have shown that, at least at a behavioral level, non-native listeners' ability to identify linguistic contrasts can be significantly improved after a short perceptual training program. We (Wang, Sereno, Jongman, & Hirsch, 2003) employed the same perceptual training paradigm to observe the cortical substrates that underlie the learning of Mandarin lexical tone. The goal was to inspect, using fMRI (functional magnetic resonance imaging), the cortical changes associated with the process of learning to identify lexical tone.

Six native speakers of American English participated. Similar to the training study, all were beginning learners of Mandarin Chinese. Subjects were evaluated prior to training and immediately after training using two identical fMRI scans (1.5 Tesla scanner). All subjects participated in a training program identical to that described earlier (Wang et al., 1999), during which subjects received feedback in correctly identifying the four Mandarin tones. Each scan (pretest scan and posttest scan) consisted of 30 images for each run: a baseline resting period of 10 images, a stimulation period of 10 images, and a baseline recovery period of 10 images. During the stimulation period, subjects performed a tone identification task in which they were required to identify 40 Mandarin words involving one of the four distinct tones in Mandarin by indicating (with a laser pointer attached to their right hand) which tone was heard. In this way, behavioral data (percent correct) for each of the participants were also collected. Two additional tasks were used to control for non-specific visual, auditory and motor components, including a task in which subjects tapped their right-hand fingers and thumb while viewing a flashing checkerboard and a task in which subjects listened to pure auditory tone sweeps. To isolate the activation that was specific to tone identification, activation of each voxel was determined by a multi-stage statistical analysis in which significant signal changes (between baseline and stimulation) were required on the two identical runs, segregating out the extraneous visual, auditory and motor activity.

We (Wang, Sereno, Jongman, & Hirsch, 2003) found that the behavioral data gathered during the scans show a very similar pattern to that obtained earlier. These data show a significant 24% increase in correct identification of Mandarin tones, from pretest accuracy levels of 64% correct to posttest accuracy levels of 88% correct. While each individual participant showed some improvement, ranging from 10% to 63% improvement, one participant's pretest accuracy scores were initially quite high (95%) and remained so after training (posttest accuracy at 95%).

Given that the behavioral data were comparable to the earlier training data, an evaluation of the cortical changes was undertaken. We found that for all participants, across pretest and posttest scans, multiple brain areas are activated bilaterally in both anterior and posterior brain regions. Table 1 lists the pretest and posttest activations as well as the conserved areas across the individual participants.

Table 1. Pre-test and post-test activation in anatomical and Brodmann's areas per subject and hemisphere, and conserved areas across five subjects

Area	Test	Subject					Conserved Areas
		RO	KD	MN	MS	KR	
LH							
GFi	Pre test	44,45,46	44,45,46	44,45,46	44,45	44,45,47	44,45
	Post test	44,45,46	44,45,46	44,45,46	44,45,47	44,45,46,47	44,45
GFd	Pre test	6	6	6	6	6	6
	Post test	6	6	6	6	6	6
GTm	Pre test	21,22,37	21	21,22,39	21,22	21,22,39	21
	Post test	21	21,22,39	21	21	21,22,39	21
GTs	Pre test	22,39,42	22	22	22	22,42	22
	Post test	22,42	22,42	22,42	22,42	22,39,42	22,42
RH							
GFi	Pre test	–	44,45,46	44,45,46	45,47	44,45,47	–
	Post test	44,45,46,47	44,45,46	44,45,46	44	44,45,47	44
GTm	Pre test	21,22	21	21,22,37	21,22	21,22,37,39	21
	Post test	21	21,39	21,37	21,22	21,22,39	21
GTs	Pre test	22,42	22	22,42	22	22,39,42	22
	Post test	22,42	22,42	22,42	22	21,22,39,42	22

LH: left hemisphere; RH: right hemisphere; GFi: inferior frontal gyrus; GFd: medial frontal gyrus; GTm: middle temporal gyrus; GTs: superior temporal gyrus.

Cortical areas consistently activated across all participants in both pretest and posttest scans were initially considered. These conserved areas include the classic areas (Broca's area and Wernicke's area) associated with language processing in the left hemisphere: the inferior frontal gyrus (BA 44 and 45) and the medial and superior temporal gyrus (BA 21 and 22). Additional consistent activation was also observed in medial frontal gyrus (BA 6, corresponding to supplemental motor cortex). Conserved activation across participants was also observed in the right hemisphere in the analog to Wernicke's area: superior temporal gyrus (BA 21 and 22). These data show that across pretest and posttest scans, multiple brain areas were activated bilaterally in both anterior and posterior regions.

Two cortical areas also show activation in posttest scans that do not show up before training. These areas include left hemisphere superior temporal gyrus (BA 42) and right hemisphere inferior frontal gyrus activation, a right hemisphere analog to Broca's area (BA 44). These findings demonstrate that new cortical areas seem to be recruited to serve the acquired tone identification function. Emergence of activation in nearby language-related cortical regions and in functionally symmetrical right hemisphere areas seems to coincide with changes at the behavioral level due to training.

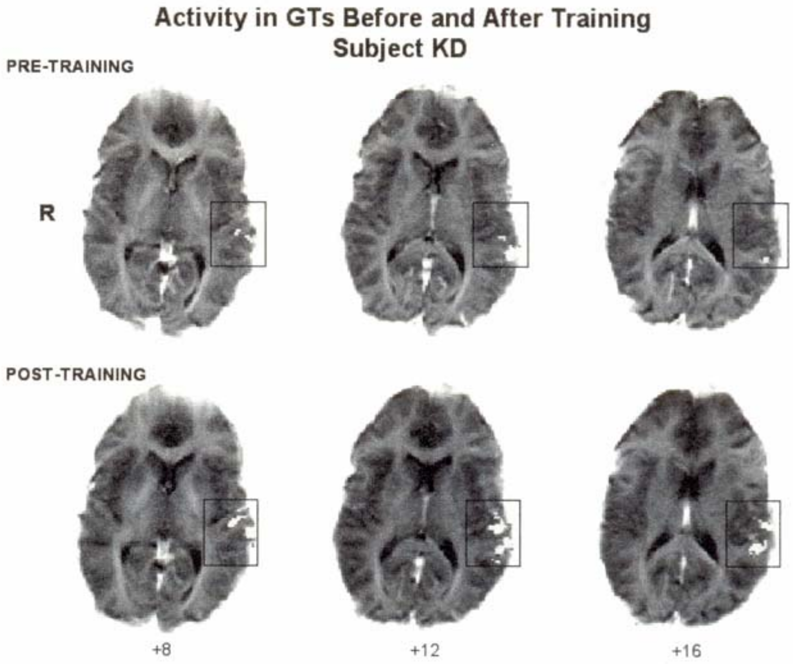


Figure 4. Pre- and post-training activity in Wernicke's Area. Activity within the left Superior Temporal Gyrus, BA 22, is overlaid on T2* base images. Representative contiguous axial slices acquired +8mm, +12mm, and +16mm superior to the AC/PC line, (subject KD), show an increase in activated volume during the lexical tone identification task after training. Voxels that passed the multistage statistical criteria within the boxed regions at a level of $p \leq 0.0005$ are white. R indicates the right side of the brain. Common activated areas in the pre- and posttest scans are not shown. (From Wang, Sereno, Jongman, & Hirsch, 2003)

To isolate additional changes from pretest to posttest, two further analyses were conducted examining location and amount of activation. First, a comparison of the centroids of the regions activated before and after training indicated that the locations on average did not differ significantly. The centroids from pretraining scans were similar in location to the posttraining scans. However, in an analysis examining amount of activation, one significant difference due to training was observed. The average number of postraining activated voxels in the left superior temporal gyrus, BA 22, exceeded the average pretraining activated voxels. This increase in activity within classically defined

Wernicke's area could not be attributed to overall activation, since overall activation, across all subjects for all areas, did not differ from pretraining to posttraining. This suggests that the increase in volume of activation in Wernicke's area was a localized effect as a consequence of tone training. An individual participant's data (Figure 4) illustrates this point, revealing an increase in magnitude and extent of posttraining activated cortex in the left superior temporal gyrus.

A final interesting observation concerns the one participant who behaviorally displayed ceiling performance at pretest and, consequently, after training, showed little change in his posttest performance (pretest: 95% correct; posttest 95% correct). While this relatively proficient learner behaviorally showed no change after training, he did exhibit some differences cortically. Although the same cortical regions were activated in pretest and posttest scans, the overall activation in these areas decreased from pretest (17,789 voxels) to posttest (6,654 voxels). While the less proficient learners showed an increase in cortical volume with training, this proficient learner became more focused. The question remains what the nature of the learning process is, with possible early cortical expansion giving way to focused cortical activation at later stages of the learning process.

Together, the fMRI data reveal cortical changes associated with lexical tone training, including an increase in volume of activation as well as the involvement of neighboring neural areas, demonstrating a cortical recruitment of resources to accomplish the task of identifying unfamiliar tones. As learning occurs, areas devoted to native language processing and contiguous cortical areas can develop specializations for non-native language function, with some indication that increased proficiency brings about progressive cortical change.

Perceptual training and tone production

While these data address the issue of perceptual learning from a behavioral and cortical standpoint, they raise the question whether this perceptual training can be transferred to production. For second language learners whose goal is to be understood, it is important to also be able to accurately produce second language distinctions. Moreover, similar gains in perception and production accuracy may lend some support to the notion that there is a unified mental representation that underlies both speech perception and speech production (e.g., Flege, 1987, 1995a; Bradlow et al., 1997), involving a reorganization of the auditory-acoustic space in second language learners. Consequently, the previous research documenting an effective procedure for training listeners to accurately perceive a novel non-native language contrast was extended to evaluate how this perceptual learning impacts production. While earlier studies have shown some support for a transfer effect of perceptual training on production for VOT contrasts and the /r/-/l/ segmental distinction using native speaker judgments as evaluation of productions (Rochet, 1995; Bradlow et al., 1997, Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999), Mandarin tone training has not been extensively considered. We (Wang, Jongman, & Sereno, 2003) therefore examined the productions of American listeners who had been trained perceptually to identify the four Mandarin tones. The participants were the 16 native speakers of American English (8 trainees and 8 controls) who had participated in the perceptual training study presented above (Wang et al., 1999). Prior to the pretest and following the posttest, all subjects were recorded reading 80 stimuli (including both previously heard and new stimuli).

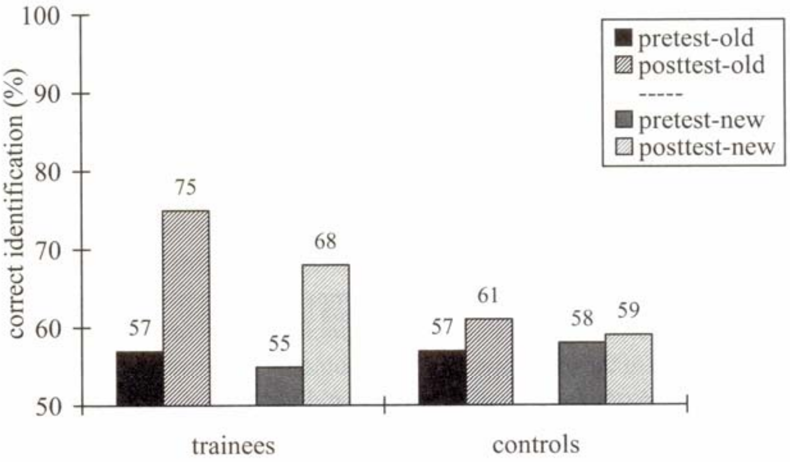


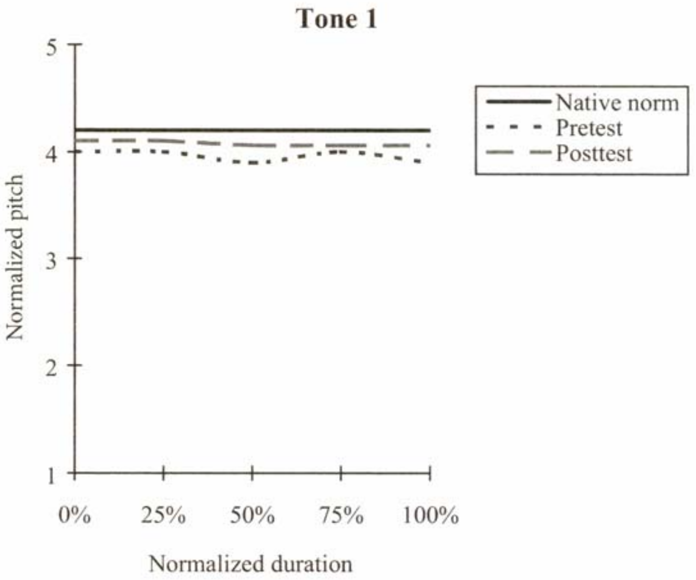
Figure 5. Mean %-correct ID of the tones from the trainees' and the controls' productions as judged by native Mandarin listeners. "Pretest-old" and "posttest-old": pre- and posttest identification of the "old" stimuli included in training; "pretest-new" and "posttest-new": pre- and posttest identification of the "new" stimuli not used in training

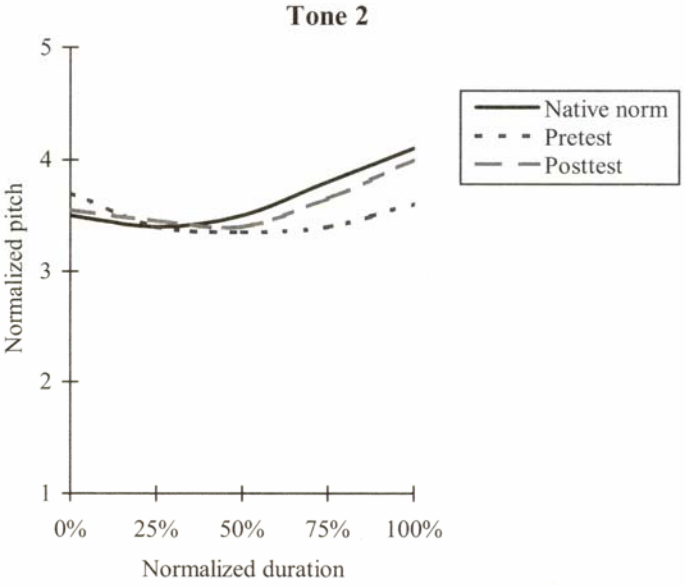
Two evaluations of the productions were undertaken: native speaker judgments and acoustic analyses. First, for the native speaker judgments, 80 native speakers of Mandarin Chinese evaluated the pretest and posttest productions by identifying the tone they heard. Figure 5 shows the mean percent correct identification of the tone productions as judged by native Mandarin listeners. Trainees showed an improvement as compared to the control subjects. Specifically, for the trainees' productions, native speaker identification accuracy scores increased from 56% in the pretest to 72% in the posttest across both the stimuli used in training (i.e., old stimuli) as well as new stimuli.

Even though there existed a large degree of variability among the eight trainees and across the four tones in terms of initial accuracy, each of the trainee's productions and each of the individual tones showed improvement after training, suggesting a consistent and robust effect of perceptual training on production accuracy. Trainees' productions result in substantial improvement in tone intelligibility as judged by native speakers of the language. Control subjects who didn't participate in the two-week training procedure showed little change in their productions as judged by native speakers, with tone identification scores from 58% in pretest to 60% in posttest (see Figure 5). Native Mandarin speakers more often correctly perceived the intended tone after participants had taken part in the training session, judging posttest productions much more accurately than those of participants who had not been trained.

A second approach to analyzing the pretest and posttest productions was to acoustically evaluate the pitch tracks and compare these productions to native speaker norms for each of the Mandarin tones. To provide native norms, four native speakers of Mandarin Chinese were asked to produce the same set of stimuli produced by the 8 trainees. Both native and non-native pitch contours were then normalized in terms of F0 to accommodate the pitch range differences among speakers and the contours were also normalized in terms of duration to adjust for differences in speaking rate (see Wang, Jongman, & Sereno, 2003 for details). For each pitch contour, F0 values were calculated at temporal points corresponding to 0%, 25%, 50%, 75%, and 100% of the contour as well as at the critical points in the contour (such as the peak, valley, rising range, and falling range) to compare overall pitch shape of native and non-native speakers. The pretest and posttest productions were then compared to the native speaker norms for each of the four tones. Figure 6 illustrates the pitch contours for each tone, comparing the pitch contours of the native speaker to the contours of the trainees' pretest and posttest productions. A comparison of the pretest and posttest productions relative to the native norm showed that the difference between the posttest values and the native norm was always smaller than that between the pretest values and the native norm, indicating a closer approximation to the native norm after training.

These results show that as a consequence of perceptual training, the posttest productions of the non-native speakers approximate the native F0 pitch contours to a greater degree than do the pretest productions. Without explicit production training, perceptual training by itself facilitates production as measured by both native speaker evaluation and acoustic analyses. It appears that perceptual learning guides production (e.g., Flege, 1997; Kuhl, 2000a, 2000b), with perceptual accuracy determining production abilities. The present production results examining the training of tonal contrasts by American learners do suggest a highly malleable speech learning system across both perception and production.


Tone 1
5
4
3
2
1
Normalized pitch
0%
25%
50%
75%
100%
Normalized duration
Native norm
Pretest
Posttest


Tone 2
5
4
3
2
1
Normalized pitch
0%
25%
50%
75%
100%
Normalized duration
Native norm
Pretest
Posttest

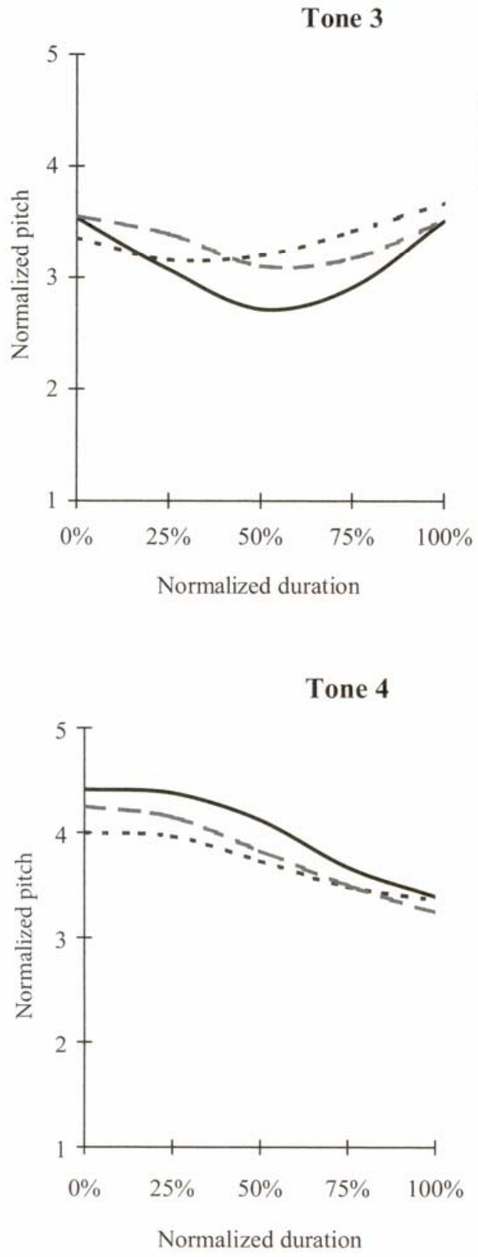


Figure 6 (a-d). Normalized pitch contours on a 5-point pitch scale, comparing the native norm to the pre- and posttest productions of the trainees, for Tones 1–4, respectively (From Wang, Jongman, & Sereno, 2003)

Conclusion

All of the above-mentioned studies document behavioral and cortical change as a result of learning a novel language contrast. The study of the acoustic and perceptual consequences of second language learning is a field almost single-handedly established by Jim Flege (e.g., Flege, 1987; Flege, 1995a; Flege & Liu, 2001; Piske, MacKay, & Flege, 2001). Many of these studies show that the adult brain obviously has the capacity to change as new phonetic contrasts are acquired. It seems that learning does not cease with the passing of a maturationally-defined critical period. The present studies also clearly show that learning non-native speech contrasts improves with training, demonstrating long-term modification of listeners' phonetic categories. Interestingly, training was accomplished using stimuli from a wide range of phonetic environments and a wide variety of talkers. While greater acoustic variability has been shown to often result in increased difficulty in identification (e.g., Mullennix, Pisoni, & Martin, 1989), the acquisition of non-native contrasts by adult second-language learners seems to benefit from exposure to greater variability during training. Phonetic categories are better acquired and longer retained when learners are exposed to these contrasts in different contexts produced by a variety of speakers. What these studies have shown is that the nature of the input is critical, with the learning of novel contrasts influenced by the nature of the training regime. While highly variable input is effective, it is also informative to recognize that not all contrasts are learned equally. Some segmental contrasts and, as we have shown, even some tonal contrasts are more difficult than others. The likelihood of category formation in a second language depends on the development of the L1 categories at the time of exposure to the L2 in combination with the perceived dissimilarity of an L2 sound from its closest L1 sound (e.g., Flege, 1987; Flege & Liu, 2001; MacKay, Flege, Piske, & Schirru, 2001). The initial phonetic system already in place is an essential determining factor. Thus, increasing second language abilities depends both on the nature of the input speakers receive and on the contribution of the first language system. It is therefore important to isolate and highlight these factors that contribute to differences between early and adult second language learners from possible differences that are the result of a neurologically defined critical period. Our findings that the adult brain retains a high degree of plasticity suggest an important role for the nature of the speech input as well as the pre-existing first language system as critical factors contributing to ultimate attainment.

A related issue is the relation between perception and production. Our finding that improvement in the perception of lexical tone is accompanied by improvement in their production even though perceptual training did not involve any production training indicates that the two are closely linked. One possibility, as suggested by Flege and others, is that perceptual learning is a prerequisite for accurate production (e.g., Flege, 2003; Flege & MacKay, 2004). If this is the case, we would not expect symmetry in the benefits gained in one domain from training in the other domain. In other words, this would predict that while training in perception improves production, an opposite pattern, namely, that training in production improves perception, is much less likely. More research is needed to explore the relation between training in both domains in detail.

Until recently, the study of the neural correlates of language has typically focused on the left hemisphere. However, learning any language involves a number of cortical regions, beyond the classically-defined Broca's and Wernicke's areas in the left hemisphere, with areas outside the traditional perisylvian language zone active as well. Moreover, right hemispheric areas must not be ignored. The selection of lexical tone as our contrast of interest was motivated by a desire to delve into the potential role of both hemispheres in processing pitch, since both the right hemisphere (overall pitch processing and emotion) and left hemisphere (linguistic tone) are involved.

One of the most interesting questions that second language research can address concerns the manner in which the brain can modify its organization over the course of the lifespan. The documented plasticity in learning is currently the topic of much research, not only in terms of language but also in other domains in both human and nonhuman animals (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Karni & Sagi, 1993; Karni, Meyer, Jezzard, Adams, Turner, & Ungerleider, 1995; Saffran, Aslin, & Newport, 1996; Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich, 1996; Ramus, Hauser, Miller, Morris, & Mehler, 2000). Although many have assumed that the brain is especially sensitive during development, researchers have only recently begun to document and appreciate how plastic the adult brain really is. While the brain was once viewed as a static organ, it is now clear that brain circuitry is constantly changing. These changes may include modifications of existing circuits or generation of new circuits. Understanding brain plasticity provides a window into constraints on acquisition of a first language as well as changes due to introducing additional languages. Experiences early in life may have quite a different effect than very similar experiences later in life for an extraordinarily complex system such as language. With second language research, we have begun to investigate the contribution of the existing language system and the modification of that system with exposure to different and novel language contrasts.