Applied Psycholinguistics **28** (2007), 565–585
Printed in the United States of America
DOI: 10.1017/S0142716407070312

Learning pitch patterns in lexical identification by native English-speaking adults

PATRICK C. M. WONG and TYLER K. PERRACHIONE
Northwestern University

Received: April 13, 2006 Accepted for publication: December 16, 2006

ADDRESS FOR CORRESPONDENCE
Patrick C. M. Wong, Speech Research Laboratory, The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: pwong@northwestern.edu

ABSTRACT
The current study investigates the learning of nonnative suprasegmental patterns for word identification. Native English-speaking adults learned to use suprasegmentals (pitch patterns) to identify a vocabulary of six English pseudosyllables superimposed with three pitch patterns (18 words). Successful learning of the vocabulary necessarily entailed learning to use pitch patterns in words. Two major facets of sound-to-word learning were investigated: could native speakers of a nontone language learn the use of pitch patterns for lexical identification, and what effect did more basic auditory ability have on learning success. We found that all subjects improved to a certain degree, although large individual differences were observed. Learning success was found to be associated with the learners' ability to perceive pitch patterns in a nonlexical context and their previous musical experience. These results suggest the importance of a phonetic–phonological–lexical continuity in adult nonnative word learning, including phonological awareness and general auditory ability.

The human nervous system has a remarkable ability to learn to integrate novel complex sounds into words. This ability is evident in both infants and adults. Although numerous studies of speech and language learning in adulthood have been conducted (e.g., Jamieson & Morosan, 1989), including recent studies examining learning-related neural changes, they were either concerned with learning foreign sounds without considering their contribution to larger linguistic contexts such as words (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1997; Golestani & Zatorre, 2004) or they focused on word learning without considering the contribution of specific phonetic features (McLaughlin, Osterhout, & Kim, 2004; Raboyeau et al., 2004). For a fuller understanding of language processing and instruction, it is essential to understand the bridge between phonemes and words, given the widely accepted notion that phonemes are the building blocks of spoken language (e.g., Chomsky & Halle, 1968). In addition, there is a mounting literature showing that infants must be able to perceive the phonetic differences between two sounds, and that they must learn that these phonetic differences have phonological importance before they can use two sound patterns contrastively for word identification (Stager & Werker, 1997; Swingley & Aslin, 2002; Werker, Fennell, Corcoran, & Stager, 2002). These results suggest a phonetic–phonological–lexical continuity for speech learning, such that more basic auditory abilities[1] (phoneme discrimination) mediate performance on higher level auditory tasks (word learning), consistent with bottom-up models of speech perception (e.g., Norris, McQueen, & Cutler, 2000). Bottom-up processes may be especially important in learning, which likely requires additional attention to acoustic details. For example, it has been shown that masking of acoustic details by noise is more detrimental to the perception of speech by less experienced (nonnative) listeners relative to more experienced (native) listeners (Van Wijngaarden, Steeneken, & Houtgast, 2002). The present study investigates two important facets of sound-to-word learning in adults: first, can native speakers of a nontone language successfully learn the use of pitch patterns for lexical identification; second, how does the phonetic–phonological–lexical continuity manifest itself in adult second language learners—specifically, what relation does more basic auditory ability have to successful nonnative language word learning.

LEARNING OF SEGMENTAL AND SUPRASEGMENTAL FEATURES

A substantial amount of recent research has expanded our understanding of how linguistic exposure early in life affects the tuning of perceptual systems for speech in adulthood (see Werker & Tees, 2005, for a review). For example, Werker and Tees (1984) found that 6- to 7-month-old infants raised in an English-speaking environment and native Hindi-speaking adults had no difficulty distinguishing the Hindi (non-English) retroflex/dental stop contrast, whereas native English-speaking adults did experience difficulty. Similarly, Polka, Colantonio, and Sundara (2002) found that English speakers improved in their perception of a /d/–/ð/ contrast during language development into adulthood, whereas French speakers showed no such improvement. These results suggest the plasticity of the perceptual systems underlying language development, such that early and extensive exposure to a phonemic contrast heightens sensitivity to that contrast specifically, shaping the perception of native and foreign speech sounds later in life (e.g., Flege, 1995). Various models have been postulated to account for the different patterns of results in adult foreign speech perception, including Best's perceptual assimilation model (e.g., Best, McRoberts, & Goodell, 2001), Flege's speech-learning model (Flege, 1995), and Kuhl's native language magnet model (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Generally speaking, these models do not focus on how nonnative phonemes can be used in words, nor do they focus on nonnative suprasegmentals (see Hallé, Chang, & Best, 2004, for an initial proposal). Despite their difficulty in perceiving some foreign sounds, it has been found that adults showed improvement in discriminating, identifying, and/or producing speech sounds that are not in their native language after short-term (weeks) behavioral training, including three-way voice onset contrasts by English speakers (e.g., Pisoni, Aslin, Perey, & Hennessy, 1982), /θ/–/ε/ contrasts by French speakers (e.g., Jamieson & Morosan, 1989), and /r/–/l/ by Japanese speakers (e.g., Bradlow et al., 1997; Lively, Logan, & Pisoni, 1993; McCandliss, Fiez, Protopapus, Conway, & McClelland, 2002).

Although almost all studies concerning phonetic learning focused on the learning of consonants and vowels (segments), the learning of suprasegmentals was also investigated. Most languages of the world use lexical tones (lexically meaningful pitch patterns) and are called tone languages (Fromkin, 2000). Mandarin Chinese is a tone language that has four lexical tones: high level (Tone 1), rising (Tone 2), dipping (Tone 3), and falling (Tone 4). For example, four different words can result when the syllable /ma/ is spoken in each of the four lexical tones. Respectively, it can mean "mother," "hemp," "horse," or "scold." Similar to the segmental learning studies discussed, Wang, Spence, Jongman, and Sereno (1999) further showed that adults were able to improve in their ability to identify Mandarin lexical tones after a short period of training. In the Wang study, native English-speaking adults were trained to identify the four Mandarin lexical tones in nonlexical contexts, that is, subjects were asked to attend to and identify the trajectory of the pitch patterns without using them to contrast word meaning. With eight sessions of training, identification accuracy increased by an average of 21%.

Taken together, these studies of speech learning show that despite years of reduced experiences with nonnative phonetic features, both segmental and suprasegmental, adults' ability to identify nonnative features can generally be improved. It is worth noting that time (e.g., 10 sessions) was often the criterion for training termination, not proficiency level. Although significant improvement was often observed, it is not clear whether all subjects learned to the fullest of their potential, and factors influencing attainment and individual differences in learning were not investigated.

SEGMENT-TO-WORD LEARNING

In the aforementioned speech-learning studies, including both segment and suprasegmental learning, the sole focus of training the adults was for them to improve in their ability to perceive the individual speech sounds or phonetic contrasts, but not to use them in actual words or larger communicative contexts. A notable exception of these adult speech training studies (including the Wang et al. study) is one conducted by Curtin, Goad, and Pater (1998). In this study, subjects learned to use three-way Thai voicing contrasts for word identification. After a fixed training period (2 days), English and French learners improved in their ability to match the auditorily presented Thai words with the correct pictures, indicating the correct use of Thai voicing for lexical purposes. However, as far as we are aware, no published studies have been conducted that examined the learnability of suprasegmentals (lexical tones in particular) for lexical identification, let alone the factors that might contribute to learning.

Although Wang et al. (1999) successfully trained native English-speaking subjects with no exposure to any tone language to identify pitch patterns embedded in various types of syllables produced by several different talkers, it remains unclear whether native English-speaking subjects can learn to use these pitch patterns for the purposes of identifying words. Anecdotal evidence suggests that even highly proficient learners of Mandarin who are native English speakers have difficulty with lexical tones, despite the fact that communication is largely successful. This is likely because of the contextual and semantic redundancy occurring in most naturalistic communicative situations. High-level attainment of Mandarin proficiency, then, may not necessarily imply a high level of accuracy in producing and/or perceiving lexical tones. This anecdotal evidence is supported by a study of Mandarin tone perception by native English-speaking adults who had several years of Mandarin instruction but whose tone perception performance was still substantially lower than that of native Mandarin speakers (Gottfried & Suiter, 1997). The primary goal of the current study was to examine whether native English-speaking adults with no previous exposure to a tone language are able to learn to use pitch patterns (suprasegmentals) in a linguistic context. In an effort to examine the subjects' individual learning ability, we used a performance-based, rather than a time-based, training termination criterion.

FACTORS ASSOCIATED WITH LANGUAGE LEARNING

In the language learning and development literature, it has been shown that phonological and morphological awareness are important indicators of learning success in Indo-European and non-Indo-European languages including Chinese (e.g., Berninger, 2001; McBride-Chang, Shu, Zhou, Wat, & Wagner, 2003). Such awareness is often defined as the learner's ability to identify and/or manipulate components of linguistic units. For example, in the commonly used Phonological Awareness Test (Robertson & Salter, 1997), subjects are asked to identify rhyming words presented in pairs in the Discrimination task of the Rhyming Subtest, and they are asked to identify one phoneme by positions in a word in the Isolation Subtest. Because these tests are generally constructed to assess Indo-European languages, no tone identification subtest is included. As far as we are aware, no studies have been conducted to examine the relationship between tone awareness and learning to use lexical tones in words in either first- or second-language learning. Given the robustness of the reported contribution of phonological and morphological awareness in language learning, including studies using formal phonological awareness test battery for assessing reading acquisition (e.g., Berninger, 2001), as well as studies suggesting that phoneme (consonant) identification (in nonstandardized test) is a prerequisite for spoken word learning (e.g., Werker & Curtin 2005; Werker et al., 2002), we expect that tone awareness will likely contribute to tone language learning. Therefore, for all our learners, a tone (pitch pattern) identification test, which addresses this aspect of phonological awareness, was administered before the onset of training, and their ability to identify these tone patterns exclusive of a linguistic context was compared with their ability to learn to use them in words.

In a recent study investigating the possible relationship between musical experience and lexical tone perception, Alexander, Wong, and Bradlow (2005) found that native English-speaking amateur musicians who had no previous exposure to any tone language showed increased ability identifying and discriminating Mandarin tones relative to their nonmusician counterparts. These amateur musicians had at least 6 years of formal private lessons in one instrument starting before 10 years old, whereas their nonmusicians counterparts had less than 3 years of musical training. However, whether these amateur musicians also possess increased ability in learning to use pitch patterns in lexical context remains unknown; this was addressed in the current study. If they indeed possess increased ability in lexical learning, that would suggest a more basic and general auditory mechanism contributes to spoken language learning, which is consistent with the general framework of bottom-up models (e.g., Norris et al., 2000) and models suggesting detailed acoustic and vocal analyses to be prerequisites of speech perception (e.g., Belin, Fecteau, & Bédard, 2004).

In the present study, native English-speaking adults who had no exposure to a tone language were trained to identify English pseudowords superimposed with three pitch patterns resembling three Mandarin lexical tones. Successful learning of the vocabulary necessarily entailed learning to use pitch patterns in words. Before training, all subjects reported their prior musical experience and participated in a separate experiment in which they were asked to identify pitch patterns similar to those superimposed in the training stimuli.

METHOD

Subjects

Subjects were 17 young adult native speakers of American English (ages = 18–26 years, mean = 20.65) who reported having no audiologic and neurologic deficits. All were undergraduate students at, or recent graduates of, Northwestern University. All but 2 subjects were right handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971); the remaining subjects were ambidextrous. None had previous exposure to a tone language at any time in life. In a self-report of foreign language experience, 13 subjects reported some experience with learning a second language ranging from "basic" to "conversational," of whom only 1 reported second-language fluency, and 4 subjects reported no exposure to a second language. To assess subjects' previous musical experience, we asked all subjects to fill out a music history questionnaire. According to self-reports, 8 subjects were amateur musicians (ages = 19–26 years, mean = 21.13), as defined by at least 6 years of formal private lessons in one instrument starting before the age of 10 (most of the subjects started earlier and had experience with multiple instruments). Nine subjects were nonmusicians (ages = 18–25 years, mean = 20.22), defined by no more than 3 years of private lessons in any combinations of instruments. Subjects were not selected specifically because of their musical background, and the proportion of amateur musicians in the study was not controlled. However, subjects who fell outside our definitions of musicianship were excluded from our study (e.g., individuals who had 6 years of musical training but started at age 11 or older). This definition of musicianship is similar to Alexander et al. (2005).

Training stimuli

The training stimuli consisted of 18 English pseudowords with pitch patterns resembling Mandarin Tones 1 (level), 2 (rising), and 4 (falling), as shown in Table 1.

Table 1. *Artificial training words*

pʰɛʃ1	dɹi1	nɛɹ1	vɛs1	nʌk1	fjut1
(glass)	(arm)	(boat)	(hat)	(brush)	(shoe)
pʰɛʃ2	dɹi2	nɛɹ2	vɛs2	nʌk2	fjut2
(pencil)	(phone)	(potato)	(tape)	(tissue)	(book)
pʰɛʃ4	dɹi4	nɛɹ4	vɛs4	nʌk4	fjut4
(table)	(cow)	(dog)	(piano)	(bus)	(knife)

Note: Subjects were trained on a vocabulary of 18 artificial words. Each word, written in the International Phonetic Alphabet, is followed by its corresponding meaning. Numbers following lexical items designate tone. Level tone is indicated by 1, rising tone by 2, and falling tone by 4, according to convention.

Tone 3, the dipping tone, was not included because it has been shown perceptually to be the most confusable tone both to native Mandarin speakers (Chuang, Hiki, Sone, & Nimura, 1972) and to second-language learners of Mandarin (Gottfried & Suiter, 1997; Kiriloff, 1969). Its exclusion should have facilitated the overall learnability of the training program. Some stimuli contain phonemes similar to those in Mandarin Chinese (e.g., Mandarin also has /p/), but all words violate Mandarin phonotactic patterns for segments such that none could be a true Mandarin word. English pseudowords were chosen because we were specifically interested in isolating the effects of learning to use pitch lexically, and it has been found that unknown words containing native phonological patterns (phonotactically legal combinations of phonemes from the native language) are easier to learn than those with nonnative phonological patterns (Feldman & Healy, 1998). As shown in Table 1, there are six sets of words with minimal pitch contrasts in each set. That is, the three words sharing the same sequence of segments differed only in terms of their lexical pitch contours. A native speaker of American English was asked to produce these words with a high pitch in a sound-attenuated chamber via a SHURE SM58 microphone onto a Pentium IV PC sampled at 44.1 kHz. These words were then resynthesized to include variants consisting of the three different pitch patterns. Pitch patterns were interpolated linearly through the voiced portion of each stimulus, using the pitch–synchronous overlap and add (PSOLA) method implemented in the software Praat (Boersman & Weeknik, 2005), which resulted in perceptually natural stimuli as judged by eight native Mandarin speakers. The starting and ending pitch points for Tone 1 were identical. This value was the mean fundamental frequency (F0) of the list of words produced by the speaker. The ending pitch point of Tone 2 was the same as Tone 1, and the starting pitch point of Tone 2 was 26% lower than its ending point. The starting pitch point of Tone 4 was 10% higher than Tone 1 and fell by 82%. These pitch contours were modeled on the values obtained by Shih (1988), and the procedures of stimulus generation were similar to Wong, Parsons, Martinez, and Diehl (2004). Other than F0, all acoustic parameters corresponded to the talker's original productions, including duration and voice quality characteristics, so that each triad of the training stimuli differed only in F0. To test the perceptual similarity of our synthesized pitch contours to

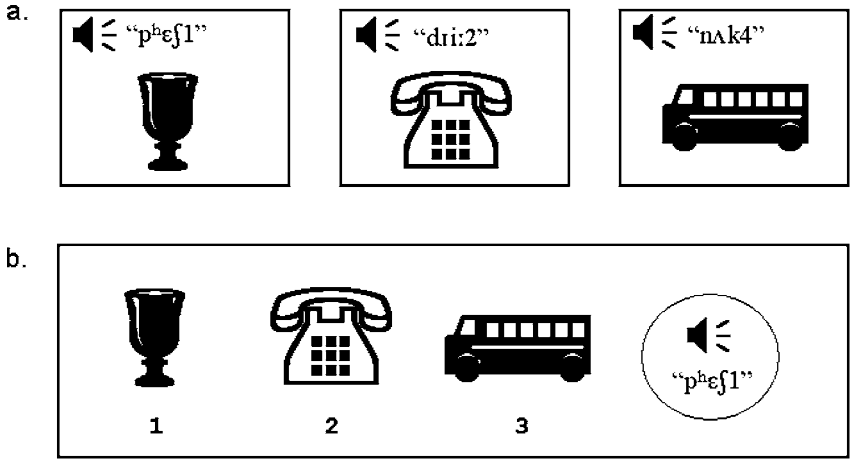


Figure 1. In a training session, subjects learned to associate the image of an object with 1 of 18 pseudowords. (a) Each word was heard four times with its corresponding picture. (b) Subjects were then quizzed over the words they just learned with feedback.

natural stimuli, eight native Mandarin-speaking individuals were asked to identify the pitch patterns (rising, falling, or level) of these training stimuli. These subjects judged our stimuli to be of the same pitch contour we intended over 97% of the time.

Training procedures

Subjects were trained to identify word meanings as depicted by drawings. As shown in Table 1, word meanings assigned to the stimuli represent high frequency English nouns (Raymer et al., 1990). Similar to Curtin et al. (1998), to facilitate learning, the 18 words were divided into six groups of three stimuli. Although each group contained all three lexical tones, no minimal pairs or triads were present within any group (e.g., /pʰɛʃ1/, /dɹi2/, and /nʌk4/).[2] The format of the training sessions is illustrated in Figure 1. In a training session, subjects learned to associate the image of an object with 1 of 18 pseudowords (a process we considered comparable to second-language word learning, because it required a mental mapping between a new [nonnative] sound sequence and a known concept); each word was heard four times with its corresponding picture (Figure 1a). Subjects were then quizzed over the words they just learned (Figure 1b). Subjects heard 1 of the words and selected the correct image from among the 3 they had just learned. During the quiz, feedback was used to help subjects recognize and correct their mistakes. At the end of each training session, subjects were presented with the 18 trained words, randomized and repeated three times (54 trials total), and were asked to identify each word by selecting the corresponding drawing out of 18 possible choices when no feedback was given. Subjects were given as much time as needed to identify the words. The score of this last word identification test of 54 trials was used to determine whether the training criterion was met. Training was terminated when subjects showed at least 95% accuracy for two consecutive sessions ("successful learning"), or when they failed to improve by at least 5% accuracy for four consecutive sessions ("less successful learning"). Because subjects were required to only learn 18 words and pilot experiments indicated that some subjects were able to achieve an accuracy level close to 100%, we established, before the onset of the actual experiment, a relatively stringent criterion (95% accuracy) to determine learning success. Subjects cannot improve more than 5% once they reached 95% accuracy or above, so we felt that our definition of successful and less successful learning was reasonable for the scope of the current study. Each session, including the training, practice quizzes, and test, lasted about 30 min. Subjects received three to four training sessions per week, with no more than one training session prescribed in a day.

Pitch pattern (tone) identification stimuli and procedures

As discussed earlier, phonological awareness was shown to be an important indicator of language learning (especially reading) success and, sometimes, disorders (e.g., Berninger, 2001). In the absence of a standard tone awareness test, we developed our own stimuli and procedures analogous to such a test. Because this test only addresses one aspect of phonological awareness, we use the specific term "pitch pattern identification test," to describe what the subjects were actually required to do.

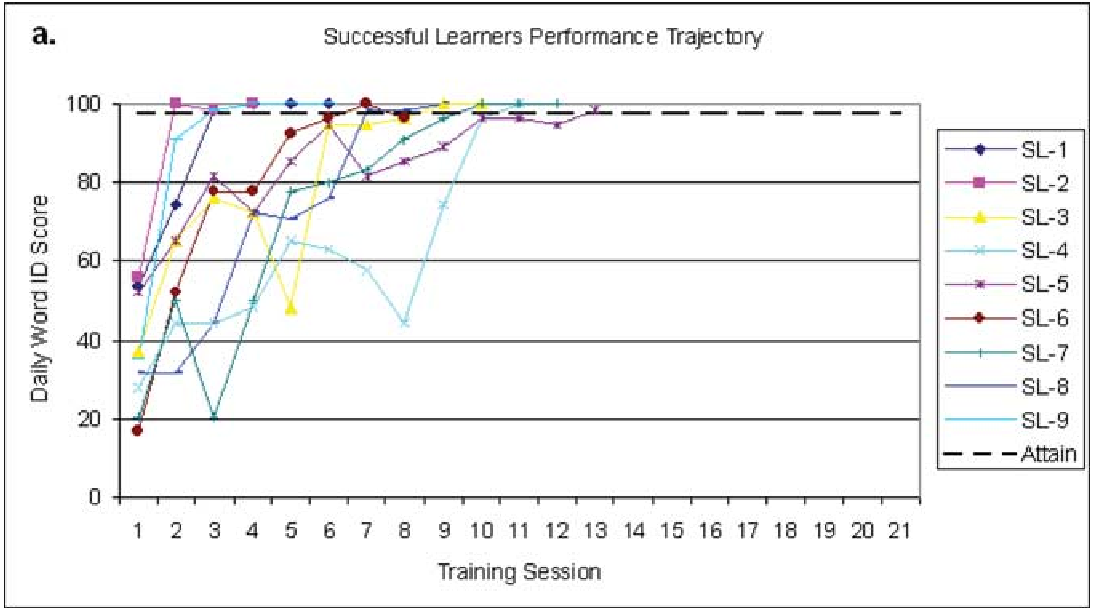
Before lexical training, all subjects participated in this pitch pattern identification test. Two male and two female speakers of Mandarin Chinese each produced five Mandarin vowels /a/, /i/, /o/, /e/, and /y/ with Mandarin Tone 1 (level tone). These vowels were then resynthesized similar to the procedures described above with Mandarin Tones 1 (level), 2 (rising), and 4 (falling) imposed. None of the speakers who produced these stimuli was the same as the native English speaker who produced the training stimuli. The set of pitch patterns for each talker was different, and their naturally produced Tone 1 was used as a reference to first guide the resynthesis of Tone 1 and subsequently Tones 2 and 4. Including three random repetitions, there were 180 trials (four speakers × five vowels × three tones × three repetitions). Subjects heard one syllable at a time and were visually presented two different pictures (i.e., → = level, ↑ = rising, and ↓ = falling), one depicting the pitch pattern of the auditory stimulus. Subjects were asked to press the response button that corresponded to the pitch pattern of the vowel they heard. For example, button A = ↑ (shown on the left side of the screen) and button B = ↓ (shown on the right side). They were familiarized with the task before proceeding with the actual experiment. Because this task was designed to test subjects' pretraining pitch perception ability, they received no feedback regarding their performance on this task, either from the computer or the experimenter. Note that this task is considered "nonlexical" because subjects were not asked to identify or access words but to only attend to the pitch patterns. These procedures resembled those in Alexander et al. (2005) comparing amateur musicians and nonmusicians who were native English-speaking adults.

RESULTS

Word-learning performance

As discussed in the Methods section, subjects' word identification (learning) performance was assessed at the end of each training session. Furthermore, no one set time determined the termination of training, but rather the training program continued until learners reached their individual asymptotic performance. Subjects' performance was considered asymptotic if they showed less than a 5% improvement for four consecutive sessions (less successful learners) or when they reached higher than 95% accuracy for two consecutive sessions (successful learner). We found that at the end of the first training session, the average (mean) word identification score for all subjects was 32.24% (range = 11.11–55.56%). At the initial point where asymptotic performance was observed ("attainment level"), subjects' average performance was at 83.23% (range = 42.60–100%), an improvement of 50.98% (see Figure 2). Nine out of the 17 subjects were successful learners and 8 were less successful learners. Figure 2 shows the performance (learning) trajectories of individual subjects separated into the two subject groups. A 2 × 2 (Group × Training) repeated measurements analysis of variance (ANOVA) revealed a main effect of training, $F\ (1, 15) = 118$, $p < .0001$, demonstrating that all subjects improved their performance. We found no significant difference between the two subject groups in the number of sessions it took to reach asymptotic performance; successful subjects as a group took 7.22 (range = 2–12) sessions to reach asymptotic performance, whereas less successful subjects took 9.38 (range = 5–18) sessions.

We also calculated the number and type of errors the subjects made. As correctly learning the new words required learning both their segmental and pitch pattern components, errors could be attributed to misidentifying one of either of these components or both. Tone-only errors were determined when the subject's response correctly matched the segments of the stimulus, but failed to match the pitch pattern (e.g., answering /pʰɛʃ1/ if the target was /pʰɛʃ4/). Subjects could also misidentify just the phonetic segments of the word (e.g., answering /pʰɛʃ2/ if the target was /dɹi2/), or both the tonal and segmental components of a syllable. At the end of the first training session, 64.23% of the errors were associated with misidentifying tone alone for all subjects. At attainment, over 97.06% of all errors were related to misidentifying tone alone. There was no reliable difference in the percentage of error between successful and less successful learners (Figure 3). A Group × Training repeated measures ANOVA revealed a main effect of training, $F\ (1, 15) = 30.483$, $p < .001$, but with no significant interaction or reliable group differences. These results showed that although learners had difficulty with both segmental and suprasegmental patterns early on, they acquired segmental contrast to a nearly perfect degree, and only pitch patterns (tonal) errors remained at attainment. This was true for both successful and less successful learners. Note that the percentage of an error type was calculated with regard to the total number errors in general. In other words, although the errors made by both groups were almost exclusively tone only toward the end, the less successful learners still made a greater quantity of errors than their successful counterparts, and these errors were pitch related. These results suggest successfully identifying pitch patterns to be the key component of learning in this training program.


a.
Successful Learners Performance Trajectory
Daily Word ID Score
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Training Session
SL-1
SL-2
SL-3
SL-4
SL-5
SL-6
SL-7
SL-8
SL-9
Attain

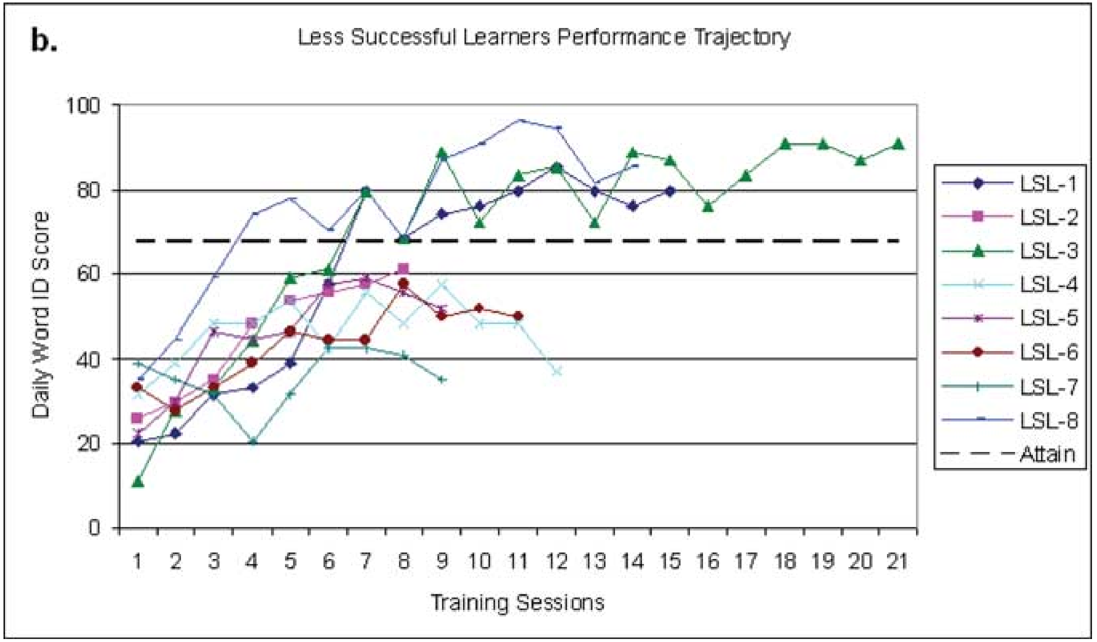


Figure 2. The trajectory of subjects' performance on the word identification task. (a) Successful learners (SL) attained word identification accuracy of above 95%, and (b) less successful learners (LSL) reached attainment at various levels below 95%. Dashed lines on each panel indicate the mean attainment of that subset of subjects. [A color version of this figure can be viewed online at www.journals.cambridge.org]

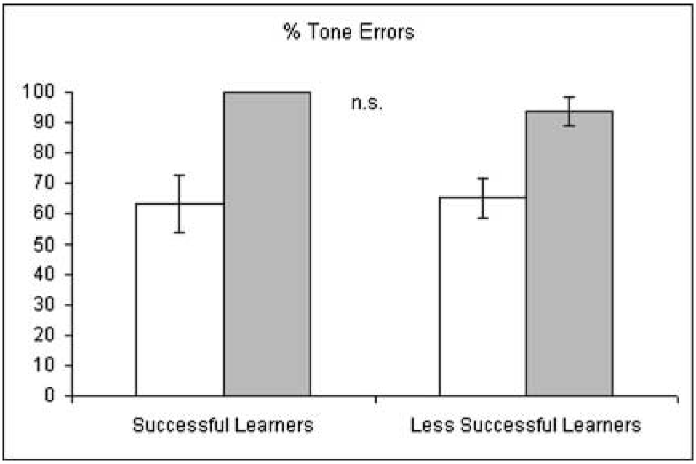


Figure 3. Successful and less successful learners did not differ significantly in the proportion of tone-only errors made during word identification either after the first session (light bars) or at attainment (shaded bars).

Nonlexical pitch pattern identification

As discussed above, before they began the training program, all subjects participated in a nonlexical pitch pattern (tone) identification task. Subjects' mean accuracy level was at 78.05% (range = 62.22–98.33%). To assess the relationship between nonlexical pitch pattern identification and word learning attainment level, a linear regression model was constructed with attainment level being the dependent variable and pitch pattern score being the predictor. We found that the pretraining pitch identification score significantly predicted the attainment level, $R^2 = 0.487$; $F\ (1, 15) = 14.22$, $p < .002$. These results are shown in Figure 4.

Further analyses of the relationship between pitch pattern identification and word learning showed that the successful learners scored significantly higher in pretraining pitch pattern identification relative to the less successful learners as a group, $F\ (1, 15) = 22.08$, $p < .0001$. In addition, we found little overlap between the ranges of pitch pattern identification scores between the two subject groups as shown in Figure 5.

Musical experience and word learning

Only subjects who fulfilled our definitions of musicianship were included in this study. We found that the distribution of amateur musicians differed significantly between the successful and less-successful learner groups, $\chi^2\ (16) = 4.50$, $p < .034$, with seven out of the eight amateur musicians (87.5%) becoming successful learners. Only two out of the nine nonmusicians (22.22%) successfully learned. In other words, seven out of our nine successful learners (77.78%) were amateur musicians. It was not surprising that we found these amateur musicians

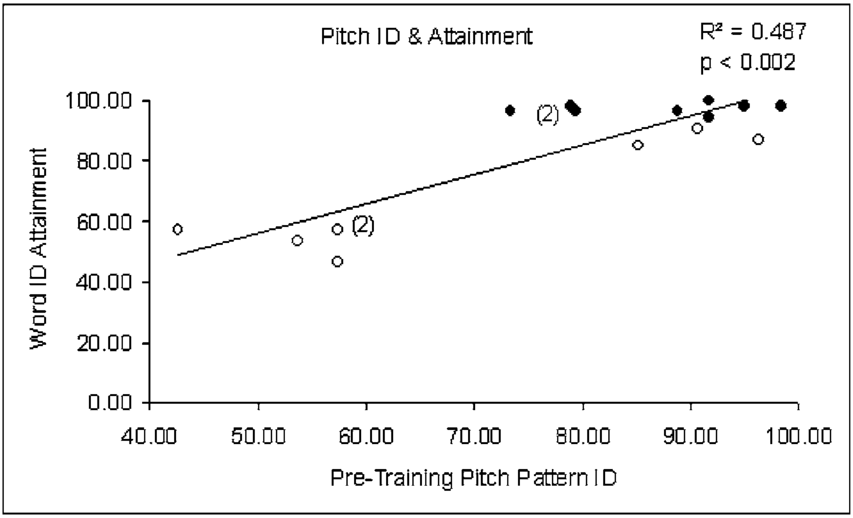


Figure 4. Subjects' pretraining pitch pattern identification score was a significant predictor of their ultimate attainment. That is, the better they were at identifying pitch patterns in a nonlexical environment, the better they were able to use those pitch patterns during word learning. Parentheses next to data points indicate the number of overlapping points. Filled and open data points represent successful and less successful learners, respectively.

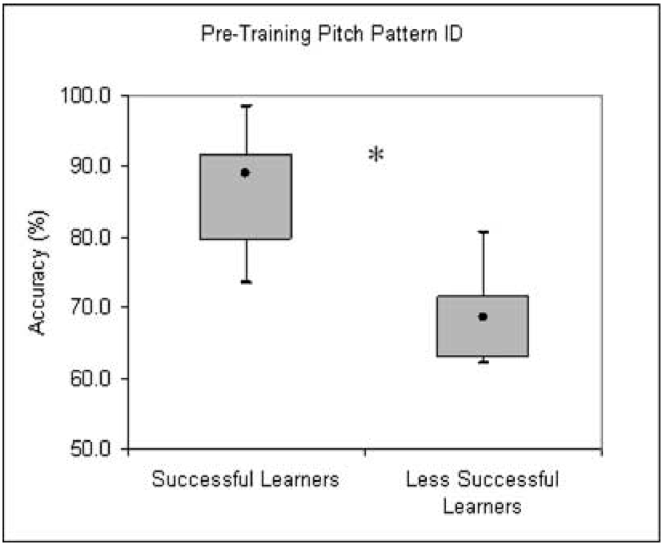


Figure 5. Successful learners and less successful learners showed little overlap in their pretraining pitch identification scores. Black points represent the group median, bars represent the maximal and minimal values, and the shaded regions represent the interquartile range.

to have increased pitch pattern identification scores relative to our nonmusicians, $F\ (1, 15) = 11.335$, $p < .004$. These results replicated those of Alexander et al. (2005) who found musicians to have increased ability in discriminating and identifying Mandarin tones relative to nonmusicians. Here, we also found that musicians tend to have an increased ability to learn to use pitch patterns in words. We also attempted to find correlations between other factors of musical experience and learning (attainment level). These factors included the length of musical training and the age of onset (when musical training started); however, none of these factors was significant. It is also worth noting that the successful learners and less successful learners did not differ appreciably in the amount of nonnative language exposure they had prior to undertaking this training program.

DISCUSSION

As far as we are aware, this is the first study demonstrating that in at least some adults, suprasegmental phonetic features that do not occur in their native language can be learned for use in words. These results build on the work by Curtin et al. (1998), who showed that individuals could learn to use nonnative segmental contrasts to identify words. Furthermore, these results are complementary to the results of Wang et al. (1999), which showed that native English-speaking subjects can improve in identifying Mandarin tones in nonlexical contexts.

Four characteristics of the current training methods and stimuli as they relate to previous studies are worth mentioning. First, the current study used a performance-based training termination, instead of a time-based criterion as used in previous speech learning studies discussed earlier. Using this method, we arguably allowed each subject to reach his/her individual learning potential and provided observations of ultimate attainment level for each individual. This method also allowed us to identify individuals who were highly successful learners and those who were not and to attempt to investigate factors that contribute to various degrees of successful learning.

Second, another characteristic of the current training method is that our training stimuli were produced by only one talker. It has been found that although speech perception (and tone perception) in multitalker conditions is more difficult for subjects, relative to single-talker conditions (Wong & Diehl, 2003), using training stimuli spoken by multiple talkers could enhance learning of the target sounds, that is, if learning is defined by generalization to stimuli that were not previously heard (e.g., Lively et al., 1993). Because the main focus of the current study is to examine the learnability of pitch patterns for lexical purposes (not stimulus–internal factors that influence learnability or generalization to new stimuli), we opted for a paradigm that was more likely to decrease learning time (i.e., be less difficult), although generalization to novel stimuli, which was not our focus, might be compromised.

Third, we chose to include resynthesized rather than natural stimuli for training, although natural stimuli have been shown to be more conducive to learning relative to fully synthesized stimuli (e.g., Logan & Pruitt, 1995). Because lexical tones are cued by multiple acoustic cues, including pitch, voice quality, and duration (e.g., Blicher, Diehl, & Cohen, 1990), we wanted to be able to conclude more definitively that at least one acoustic factor (pitch) was attended to by the learners, and failure to learn was not necessarily related to an inability to integrate multiple acoustic factors. It is worth emphasizing that the resynthesized stimuli using the PSOLA method as stated above, including our training stimuli, generated highly natural sounding stimuli as judged by native Mandarin speakers. We believe, therefore, that a reduction of naturalness was not a concern. However, as a follow-up study, one may seek to assess if natural stimuli could enhance or reduce learning, including the consideration of various acoustic cues (duration, voice quality, etc.) that are potentially important in lexical tone perception by native speakers. Furthermore, because we have established in the current study that some adults are able to learn to use pitch patterns in word identification, we are currently investigating the differences between training programs that include training stimuli from one versus multiple talkers in terms of the length of training as well as the degree of generalization to untrained stimuli.

Fourth, a characteristic of the current study is that we used English pseudowords rather than Mandarin syllables as training stimuli. As discussed earlier, we were sensitive to the fact that nonnative phonetic (segmental) sequences could be difficult to learn (Feldman & Healy, 1998). By using English pseudowords spoken by an English speaker, we eliminated the training of nonnative segments or accented syllables and instead required our subjects to focus on learning only a nonnative suprasegmental (pitch) contrast. Our research question did not concern the Mandarin language per se, but whether a phonetic feature (viz., lexical pitch pattern) that is used in most languages of the world (including Mandarin) can be learned by nontone language speakers. Therefore, we believe it was reasonable to utilize English segmental patterns.

Tone awareness and word learning

We found in the present study that the ability to identify pitch pattern in a nonlexical context significantly predicted learners' ability to use pitch in newly learned words, as indicated by a regression model, and comparisons may be made between pitch pattern identification scores from successful learners and less successful learners in terms of group averages and the minimal overlap of ranges of scores. These results concerning the connection between pitch identification (one aspect of tone awareness) and word learning involving pitch is reflective of the broader language learning literature showing metalinguistic awareness, including phonological and morphological awareness, to be important language learning indicators (e.g., Berninger, 2001; McBride-Chang et al., 2003). Our results show that such metalinguistic (phonological) awareness is restricted not only to an alphabetic system or segmental phonemes but also to suprasegmentals.

It is important to point out that we do not see pitch identification to be the only factor influencing word learning. We found in our regression model that about 50% of the variance in attainment level was explained by pitch identification. We do not doubt that other factors contribute to successful sound-to-word learning. Just as a variety of cognitive and environmental factors such as phonological working memory affect the development of native language (Adams & Gathercole, 1996), it has also been found that multiple factors contribute to successful second language learning and, to a certain extent, nativelike competency, analogous to our highly successful learner group (see Robinson, 2005, for a review). Some of these additional factors include verbal working memory (e.g., Miyake & Friedman, 1998), motivation, and age of onset, as well as length, intensity, and quality of training (e.g., Birdsong, 1999; Bongaerts, 1999). However, with the ability to explain half of the variance, we believe nonlinguistic pitch identification (phonetic awareness) to be a highly important factor for word learning involving such an acoustic cue. As discussed below, the fact that by the end of training subjects were making almost exclusively pitch-related errors regardless of attainment level further lends weight to the specific importance of pitch perception ability for learning here.

It is also noteworthy that both successful and less successful learners made the same type of errors. Namely, they were both able to learn segments to a near perfect competency and the errors they made toward the end of training were almost entirely related to pitch, although the less successful learners made greater overall numbers of pitch errors than successful learners. This appears to suggest that the less successful learners did not have decreased working memory associated with segmental patterns relative to the successful learners. (For the importance of phonological working memory in relation to second-language learning, see Baddeley, Gathercole, & Papagno, 1998; Cheung, 1996; and Speciale, Ellis, & Bywater, 2004.) At least their memory is sufficient to perform well in our learning program if pitch were never involved.

Music and tone

An interesting aspect of our results is that subjects in the successful learners group tended to have increased musical experience. Music and language can both be described in terms of many common structural characteristics, such as syntactic, generative, combinatorial, and rhythmic features (Patel, 2003). These structural characteristics are supported by organs of the peripheral and central nervous systems, including muscles of the vocal folds and vocal tract, the cochlea, as well as the neocortex, including Broca's area and the auditory cortex (Denes & Pinson, 1993; Wong et al., 2004). It has been shown that musicians displayed various neuroanatomic and neurophysiologic differences relative to nonmusicians. For example, at least some musicians (those with perfect pitch) showed left-biased asymmetry in the size of planum temporale (Keenan, Thangaraj, Halpern, & Schlaug, 2001; Schlaug, Järcke, Huang, & Steinmetz, 1995). Musicians also showed increased cortical representations in the auditory cortex, as reviewed by listening to piano tone relative to pure tone of the same frequencies (Pantev et al., 1998). These neuroanatomic and neurophysiologic differences, especially those around the auditory cortex, have been implicated in explaining why musicians have better verbal memory (Brandler & Rammsayer, 2003; Chan, Ho, & Cheung, 1998). Given that verbal working memory has been found to correlate with second language learning (Miyake & Friedman, 1998), it is not unreasonable to expect that musicians might have increased ability in learning a second language. Although research has shown that musical training improved spatial–temporal reasoning and math skills in children (Graziano, Peterson, & Shaw, 1999), relatively little attention has been drawn to the impact of musical training on second-language word learning and how specific musical abilities may contribute to phonetic perception and word learning. The current study provided some initial evidence for future studies of the transferability of musical ability and language learning that specifically pertains to sound-to-word learning. It is possible that this line of research may provide more specific information with regard to if or why musicians are more proficient in foreign language comprehension (Buck & Axtell, 1986). In addition, we acknowledge that the present study focused only on the relation of musicianship to learning a specific nonnative suprasegmental phonological contrast. It remains to be shown whether musicianship likewise facilitates learning segmental contrasts or other aspects of language.

We are cautioned by the fact that our definition of musicianship is somewhat arbitrary. Without the guidance of existing music literature concerning an appropriate definition that we are aware, we chose to continue using a definition similar to what has been found in the literature (Alexander et al., 2005), which includes consideration of age of onset and years of training, similar to language learning studies (Birdsong, 1999). Furthermore, at this point, we are not certain whether our successful learners possessed an increased pitch perception ability that is more general, or if their ability was more specific to distinguishing level, rising, and falling pitch patterns as used in the current study. It is also not certain whether the successful learners have an increased general auditory processing and memory ability, and their increased pitch ability is only a by-product. Moreover, we are aware that these results do not explain whether innate ability or increased experience with pitch owing to musical training facilitated word learning. However, we do believe that the current study provided some insights into the relationship between music and language processing and learning. Advocates of a modular view of music processing (Peretz & Coltheart, 2003) support the idea of shared resources for music and speech at the level of early acoustic analysis, but maintain that detailed spectral and temporal processing is specific to either speech or music. If there is greater transferability between music and language, it can support the idea that even if distinct modules exist for speech and music, the level of divergence may be located beyond initial acoustic analysis and extend to the formation of memories and associations of sound. Specifically, the experience of perceiving and learning one category of sounds (music) may influence, or perhaps facilitate, the perception and learning of another category of sounds (speech).

Future directions

Several interesting questions remain. It has been shown that long-term exposure (native language experiences) to the phonological patterns of tone languages can significantly affect the perception of pitch patterns (Bent, Bradlow, & Wright 2006); however, it is not known if short-term exposure could result in the same effects. The extent of such plasticity could be assessed by testing learners before and after short-term training on tasks similar to the Bent et al. (2006) study. Another important remaining question is how to effectively train the less successful learners of the current study to raise their word identification performance closer to that of the successful trainees. It is most likely that the current training program is not conducive to optimal learning for certain individuals, especially those with a poorer pretraining pitch identification ability. If nonlexical pitch identification before training is indeed associated with lexical training, as suggested by the current results, it is conceivable to first prescribe a nonlexical pitch training program similar to Wang et al. (1999) to bring nonlexical pitch identification to a higher level prior to lexical training. These questions involving different training methods, including whether stimuli should be grouped to highlight the phonetic feature in question, are being addressed in our ongoing research. In addition, we are conducting experiments investigating the neural correlates of sound-to-word learning (e.g., Wong et al., in press; Wong, Perrachione, & Parrish, in press), and voice learning (Perrachione & Wong, 2007) to complement existing studies on cross-linguistic lexical tone perception (Wong et al., 2004) and lexical tone learning in nonlexical contexts (Wang, Sereno, Jongman, & Hirsch, 2003). It is our hope that this line of research will provide evidence for the most effective second language pedagogy.

ACKNOWLEDGMENTS

The authors thank Jay Mittal, Carson Lam, Ann Bradlow, Diane Arcuri, Gnyan Patel, Arth Srivastava, and Purav Jesrani for their assistance in this research. This work is supported by Northwestern University and the National Institutes of Health Grants HD051827 and DC007468 (to P.W.).

NOTES

1. Throughout the text, we use the term "basic" auditory ability to refer to the relative level of acoustic, phonetic, and linguistic analyses. For example, perceiving pitch contours is a more "basic" ability than the use of pitch in word learning.
2. It is outside the scope of the current study to compare detailed methodological differences of our current basic training protocol. However, it is worth mentioning that we are currently conducting experiments comparing how stimulus grouping may affect learning; specifically, we are putting stimuli in groups of minimal triads to see whether learning becomes more efficient.

REFERENCES

Adams, A.-M., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology, 49A*, 216–233.

Alexander, J., Wong, P. C. M., & Bradlow, A. (2005). *Lexical tone perception in musicians and nonmusicians.* Paper presented at the Interspeech 2005 (Eurospeech) 9th European Conference on Speech Communication and Technology, Lisbon, September 2005.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 3*, 129–135.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 97–103.

Berninger, V. (2001). *Process assessment of the learner: Test battery for reading and writing.* San Antonio, TX: Psychological Corporation.

Best, C., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*, 775–794.

Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1–22). Mahwah, NJ: Erlbaum.

Blicher, D. L., Diehl, R., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 37–49.

Boersma, P., & Weeknik, D. (2005). Praat: Doing phonetics by computer (v. 4.3.04). Retrieved from http://www.fon.hum.uva.nl/praat/

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Erlbaum.

Bradlow, A., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977–985.

Brandler, S., & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and nonmusicians. *Psychology of Music, 31*, 123–138.

Buck, R. M., & Axtell, T. L. (1986). Music & foreign languages: Complementary fields of endeavor. *The Language Teacher, 10*, 4–7.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature, 396*, 128.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867–873.

Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1972). *The acoustical features and perceptual cues of the four tones of standard colloquial Chinese.* Paper presented at the Seventh International Congress on Acoustics, Budapest.

Chomsky, N., & Morris, H. (1968). *The sound pattern of English.* New York: Harper & Row.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research, 14*, 389–405.

Denes, P. B., & Pinson, E. N. (1993). *The speech chain: The physics and biology of spoken language.* New York: W. H. Freeman.

Feldman, A., & Healy, A. F. (1998). Effect of first language phonological configuration on lexical acquisition in a second language. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psychological studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Flege, J. E. (1995). *Second language speech learning: Theory, findings, and problems*. Baltimore, MD: York Press.

Fromkin, V. A. (2000). *Linguistics: An introduction to linguistic theory.* Oxford: Blackwell.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage, 21*, 494–506.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics, 25*, 207–231.

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal training. *Neurological Research, 21*, 139–152.

Hallé, P. A., Chang, Y.-C, & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*, 395–421.

Jamieson, D. G., & Morosan, D. E. (1989). Training non-native speech contrasts in adults: Acquisition of the English /θ/–/ð/ by francophones. *Perception and Psychophysics, 40*, 205–215.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *NeuroImage, 14*, 1402–1408.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 20*, 63–67.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351–377). Baltimore, MD: York Press.

McBride-Chang, C., Shu, H., Zhou, A., Wat, C.-P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology, 95,* 743–751.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). 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, 2*, 89–108.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*, 703–704.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as "language aptitude." In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psychological studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–370.

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

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811–814.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*, 674–681.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691.

Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the organization of human auditory cortex. *Neuropsychologia, 45*, 1899–1910.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 297–314.

Polka, L., Colantonio, C., & Sundara, M. (2002). A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109*, 2190–2201.

Raboyeau, G., Marie, N., Balduyck, S., Gros, H., Demonet, J. F., & Cardebat, D. (2004). Lexical learning of the English language: A PET study in healthy French subjects. *NeuroImage, 22*, 1808–1818.

Raymer, A. M., Maher, L. M., Greenwald, M. L., Morris, M. K., Rothi, L. J. G., & Heilman, K. M. (1990). *The Florida Semantics Battery*. Unpublished manuscript.

Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America, 106*, 1555–1565.

Robertson, C., & Salter, W. (1997). *The Phonological Awareness Test.* East Moline, IL: LinguiSystems, Inc.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46–73.

Schlaug, G., Jäncke, L., Huang, Y. X., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*, 699–701.

Shih, C. L. (1988). Tone and intonation in Mandarin. *Work Papers of the Cornell Phonetic Laboratory, 3*, 83–109.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics, 25*, 293–321.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484.

Van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111*, 1906–1916.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience, 15*, 1019–1027.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649–3658.

Werker, J. F., Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*, 197–234.

Werker, J., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1–30.

Werker, J., & Tees, D. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866–1878.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding. In *Plasticity and commitment in language systems of the brain.* New York: Wiley.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization of inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46*, 413–421.

Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience, 24*, 9153–9160.

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (in press). Neural characteristics of successful and less-successful speech and word learning in adults. *Human Brain Mapping*.

Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., et al. (in press). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex*.