Chapter 4
Individual Differences in Lexical Tone Learning

Erin M. Ingvalson and Patrick C. M. Wong

Abstract It is now well established that second language learning training results in large individual variation in learning outcomes. Native English speakers learning the lexical tones of Mandarin Chinese are no exception (e.g., Wang et al., 1999). In this chapter, we review a series of studies undertaken by our group investigating both the sources of individual differences in lexical tone learning (Chandrasekaran et al., 2010; Wong et al., 2007, 2008) and how such differences can be mediated by matching learners to the training paradigm that is best suited to their baseline phonological perception (Ingvalson et al., 2013; Perrachione et al., 2011). We include studies that have sought to identify possible genetic markers of individual variation in lexical tone learning outcomes (Wong et al., 2012a, b) and expansion of our training paradigms to older adults (Ingvalson et al., 2017) to provide further insight into individual variation across learning populations.

In second language speech perception research, there have been numerous efforts to train listeners to perceive non-native speech segments (e.g., Ingvalson, Holt, & McClelland, 2012; Iverson & Evans, 2009; Logan, Lively, & Pisoni, 1991; Sebastian-Galles & Soto-Faraco, 1999). More recently, these efforts have turned to training listeners to perceive lexical tone, recognizing that though lexical tone is a suprasegmental contrast, it nonetheless represents a novel phonological dimension to listeners who are unfamiliar with tone languages (e.g., Shen, 1989). As with non-native segmental training, non-native lexical tone-training research has found extensive individual variation in outcomes. We first review the history of lexical tone training,

E. M. Ingvalson
School of Communication Science and Disorders, Florida State University, Tallahassee, USA

P. C. M. Wong (✉)
Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, S.A.R., China
e-mail: p.wong@cuhk.edu.hk

Brain and Mind Institute, The Chinese University of Hong Kong, Room G03, Leng Kau Kui Building, Shatin, N.T., Hong Kong, S.A.R., China

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*, Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-15-7606-5_4

demonstrating that though traditional approaches to non-native speech perception training are successful in aggregate, there remain extensive individual differences in learning outcomes, with some learners approaching native-like performance post-training whereas other learners make very little progress over the course of training. We then seek to identify the sources of individual differences in learning outcomes, looking at such disparate possibilities as musical training, neuroanatomical and neurophysiological variability, and genetic variation. Having reviewed the possible sources of individual variation, we review work that seeks to optimize learning outcomes. We close this chapter by highlighting our groups' ongoing efforts into lexical tone training, branching out into understanding learning differences between older versus younger adults and seeking to understand the mechanisms of pitch perception and pitch learning in individuals with congenital amusia.

4.1 History of Lexical Tone Training

Wang, Spence, Jongman, and Sereno (1999) undertook among the first efforts to train native English listeners to perceive lexical tone. Building on the high-variability phonetic training paradigm developed by Logan et al. (1991) to train native Japanese listeners to perceive the English /r–l/ contrast, they trained native English listeners to perceive the four tones of Mandarin Chinese (we leave a discussion of the tone contours of Mandarin for elsewhere in this volume; for one example see Lee & Wiener, Chap. 3). This paradigm had been shown to lead to successful generalization to both untrained tokens produced by trained talkers as well as untrained tokens produced by untrained talkers (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Logan et al., 1991; but see Ingvalson, Ettlinger, & Wong, 2014 for potential limitations with the paradigm). Consistent with earlier uses of the paradigm, following training, trained listeners were more accurate at identifying both trained and novel tokens and talkers relative to untrained listeners. Perhaps more impressively, the degree of tone confusion decreased substantially following training. However, there were extensive individual differences in learning outcomes, with some trained listeners making large gains in their tone identification performance whereas other trained listeners made much smaller gains.

Building on the initial perceptual training study, Wang et al. (2003a) went on to determine if perceptual training would lead to improvements in learners' ability to produce Mandarin tones (c.f. Bradlow et al., 1997). Both perceptual judgments by native Mandarin listeners and acoustical measurements taken pre- and post-perceptual training demonstrated that perceptual training alone was sufficient to improve native English speaker production of Mandarin tones. Like in the perceptual training, though, there were extensive individual differences in outcomes, with some learners making large gains in perceptual accuracy (both as judged by the native listeners and in the acoustic measurements) whereas other listeners made much smaller gains.

Finally, Wang et al. (2003b) investigated the cortical effects of high-variability phonetic training of Mandarin tones in native English listeners. Using a functional magnetic resonance imaging (fMRI) paradigm during which listeners identified lexical tones, they determined that, pre-training, all listeners showed activation in areas associated with language processing, including Broca's area, Wernicke's area, auditory cortex, and supplementary motor areas. Post-training, for all listeners, improvements in lexical tone identification were associated with an increase in the spatial extent of activation on left superior temporal gyrus and the emergence of activity in right inferior frontal gyrus (see also Yu, Wang, & Li, Chap. 5, this volume). This pattern of activation was taken as evidence that learning a novel phonology involves expansion of existing language-related areas and recruitment of additional areas. Once again, though, there were individual differences in learning gains and the fact that these individual differences in learning gains are correlated with individual differences in patterns of neural activation suggests that the source for these individual differences in learning success may lie in the way the brain is processing lexical tone. We explore this possibility more fully in the following section.

There have been other training efforts to teach native English listeners to perceive lexical tone beyond using the high-variability phonetic training paradigm for native English listeners learning Mandarin tone (e.g. Wang et al. 1999, 2003a, 2003b). Caldwell-Harris, Lancaster, Ladd, Dediu, and Christiansen (2015) sought to determine first, whether lexical tone could be learned in a statistical learning paradigm (Saffran, Aslin, & Newport, 1996), and, second, what learner-level characteristics predicted the extent to which listeners were able to learn lexical tone in the statistical learning paradigm. The training stimuli were based on the lexical tone patterns of African languages. Using African lexical tone patterns as stimuli allowed native speakers of Asian tone languages—including Mandarin Chinese and Thai—to participate, permitting investigations on the role of existing lexical tone knowledge on the statistical learning on novel lexical tone patterns. Following the exposure phase, listeners were presented with minimal pairs that differed either segmentally or suprasegmentally and asked to indicate which member of the pair was a "word" in the earlier speech stream (see Saffran et al., 1996 for a fuller description of the statistical learning paradigm). Both listeners familiar with a tone language and those who were not (i.e., native English listeners) performed equally well-identifying words using segmental cues. However, those listeners who spoke a tone language in addition to English were significantly better at identifying words using lexical tone cues than listeners who did not speak a tone language; listeners who had no knowledge of a tone language were at chance on this task. Though some earlier work suggested that musical training influenced listeners' ability to categorize lexical tone (Alexander, Wong, & Bradlow, 2005; Chang, Hedberg, & Wang, 2016; Kempe, Bublitz, & Brooks, 2015; Wu et al., 2015), regression analyses indicated no effect of musicianship, instead indicating that Asian language experience was the strongest predictor of success learning tone words.

Another investigation in lexical tone training also utilized a statistical learning paradigm. Building on the unimodal and bimodal distributions developed by Maye and colleagues (Maye, Werker, & Gerken, 2002), the hypothesis tested was whether passive listening to a speech stream that included more exemplars from the endpoints of the lexical tone continuum—a bimodal distribution—would result in better learning of the lexical tone contrast than passive listening to a speech stream that included more exemplars from the center of the lexical tone continuum—a unimodal distribution (Ong, Burnham, & Escudero, 2015). Passive listening resulted in no learning in either condition, but a second experiment that required listeners to actively attend to each sound did produce better learning in the bimodal condition relative to the unimodal condition. Though statistical learning paradigms are meant to capture naturalistic learning, these two studies (Caldwell-Harris et al., 2015; Ong et al., 2015) suggest that to the extent second language learners are able to learn lexical tone via statistical learning, and it is very limited and may only be successful for those listeners who are already proficient in a language that utilizes tone phonemically. As a result, for the remainder of this chapter we will focus on individual difference in learning using variations of the high-variability phonetic training paradigm, and how this paradigm might be modified to optimize individual outcomes.

4.2 Individual Differences in Lexical Tone Learning Outcomes

Discussed above, though training native English listeners to perceive Mandarin Chinese lexical tone led to improved identification of trained talkers and tokens, improved identification of untrained talkers and tokens, improved perception, and changes in patterns of neural activation (Wang et al., 1999, 2003a, 2003b), there were extensive individual differences in learning outcomes. In segmental speech learning, both at the first and second language level, sensitivity to the to-be-learned phonology had been found to be important for successful learning (e.g., Werker & Stager, 2000), leading to the hypothesis that a similar sensitivity to tone may be important for successful lexical tone learning (Wong & Perrachione, 2007). Noting also that monolingual English-speaking musicians had performed better than monolingual English-speaking non-musicians when identifying lexical tone (Alexander et al., 2005), Wong and Perrachione (2007) further hypothesized that musical training may be important for lexical tone acquisition (note that this is in the context of the high-variability phonetic training paradigm and the lack of a music effect in Caldwell-Harris et al. (2015) is likely due to the limitations of the statistical learning paradigm for lexical tone). Though they used many of the hallmarks of the high-variability phonetic training paradigm—natural speech tokens, multiple talkers, multiple phonetic contexts, and feedback—instead of training listeners to identify lexical tones (c.f. Wang et al., 1999), Wong and Perrachione utilized a lexical learning task in which listeners were trained to associate each tone–syllable pair with a meaning. Importantly, each syllable was matched with three possible tones, meaning that listeners had to master the lexical tone contrasts in order to successfully complete the word-learning task. Listeners were trained to criterion, defined as 95%

accuracy for two consecutive sessions (successful learners) or as a failure to improve by more than 5% over four consecutive sessions (less successful learners). Consistent with the hypothesis that phonological sensitivity for lexical tone is an important predictor of learning success, a pretest that measured listeners' ability to identify pitch patterns accounted for 49% of the variance in listeners' learning outcomes. Also as expected, musicians were more likely to be successful learners than were non-musicians (78% of successful learners were musicians, whereas only 12% of less successful learners were musicians).

In a similar paradigm, Cooper and Wang (2012) trained native English listeners and native Thai listeners to learn words that differed on the basis of Cantonese tones. For the native English listeners, final learning performance was reliably predicted by baseline ability to identify Cantonese tones, accounting for 59% of the variance (Wong & Perrachione, 2007); conversely, baseline tone identification performance was not a reliable predictor of final learning outcomes for native Thai listeners. There was also a significant benefit of previous musical experience for native English listeners, with musicians having a higher level of tone word learning than non-musicians. However, musicianship was not found to be beneficial for the native Thai listeners, in that there was no difference in the ultimate learning performance between the musicians and non-musicians. Finally, there was a significant interaction of language background and musicianship such that native English musicians' final learning performance was on par with that of native Thai listeners. Though there is a clear benefit to earlier experience with a tonal language in learning a new language that uses lexical tone (c.f., Caldwell-Harris et al., 2015), these data, together with the data from Wong and Perrachione (2007), paint a preliminary picture of the possible sources of individual differences in lexical tone learning in listeners who are not native speakers of a tone-based language: baseline aptitude for identifying pitch patterns, possibly as a result of previous musical training.

Wang et al. (2003b) found correlations between listeners' behavioral outcomes and changes in the patterns of neural activation following training to identify lexical tone, demonstrating individual variation in both the post-training identification performance and the post-training patterns of neural activation. Wong, Perrachione, and Parrish (2007) hypothesized that such individualized variation in patterns of neural activation could differentiate successful versus less successful learners; in particular, because the training task of Wong and Perrachione (2007) required listeners to associate syllable–tone pairs with lexical meanings, they hypothesized that successful learners would show post-training activation in a relatively small network consisting primarily of regions associated with language processing whereas less successful learning would show post-training activation in a more diffuse network that incorporated areas associated with attention and nonlinguistic pitch processing. The participants from Wong and Perrachione participated in an fMRI protocol that asked them to discriminate the pitch patterns in the training words both before and after training. The data were consistent with the hypotheses, and successful learners showed a shift toward a language-based network of activation post-training relative to less successful learners, where post-training activation was more diffuse (see also Yu, Wang, & Li, Chap. 5, this volume). Thus, not only does baseline aptitude for pitch perception influence behavioral outcomes, it also influences the pattern of neural activation post-training. Perhaps more interestingly, there were differences in the patterns of neural activation between the successful and less successful learners in response to the lexical tone stimuli pre-training in regions associated with language processing and these differences in pre-training activation reliably predicted which listeners would ultimately master the lexical tone task. We are therefore seeing evidence that the differences in learners' post-training outcomes learning lexical tone can be traced to pre-training differences in the way listeners' brains are responding to the training stimuli.

We have now seen evidence that behavioral responses to pitch patterns (tones removed from their lexical context) and neural patterns of activations in response to lexical tone both reliably predict which learners will successfully learn words that differ in lexical tone. Completing this series, Wong et al. (2008) investigated whether there were neuroanatomical differences between successful and less successful learners that could reliably predict learning success. Relative to less successful learners, successful learners had a larger volume of Heschl's gyrus; Heschl's gyrus has been associated with nonlinguistic pitch processing and language learning and had therefore been hypothesized to be important for learning of lexical tone languages. Across this series of three studies, then, we see evidence that individual variation in lexical tone learning can be traced back to individual variation in neuroanatomy and neurophysiology and that these individual differences in neuroanatomy and neurophysiology can be reliably predicted via a behavioral test for listeners' baseline aptitude for phonological sensitivity to pitch patterns.

The above neuroanatomical and neurophysiological results, as well as the original individual differences in neural responses collected by Wang et al. (2003b), were all collected from the cortex. However, there are structures in the midbrain known to be important for auditory processing, and differences in the ability to perceive pitch patterns at the level of these structures could also be predictive of variation in lexical tone learning. One such possible structure is the inferior colliculus (IC), identified as a midbrain structure involved in auditory processes. To test the possibility that individual variation in IC activity might predict lexical tone learning success, before and after training native English-speaking listeners completed an fMRI paradigm in which they listened to repeated instances of the lexical tone stimuli used in training (similar to the stimuli from Wong & Perrachione, 2007) as well as an auditory brainstem response (ABR) task in which they listened to a single syllable with a rising tone (Chandrasekaran, Kraus, & Wong, 2011). Training procedures were similar to Wong and Perrachione (2007), except that all listeners trained for nine days, rather than to criterion. Listeners who showed a reduction in neural response in the IC following repeated instances of the same lexical tone—an indicator of more efficient neural processing, called repetition suppression—also showed better tracking of the rising pitch on the ABR measure. Conversely, listeners who showed an increase in neural response in the IC following repeated instances of the same lexical tone—an indicator of less efficient neural processing, called repetition enhancement—also showed poorer tracking of the rising pitch on the ABR measure. Those listeners who showed a repetition suppression response pre-training were also those listeners who went on to be successful learners, demonstrating that neurophysiological differences between successful and less successful learners exist not only at the cortical level, but also in midbrain structures important for auditory processing.

With the exception of the neuroanatomical measurements of Heschl's gyrus, all the above studies asked the participants to engage with the lexical tone stimuli in some way, even if by passive listening to repeated instances. However, the brain engages in spontaneous activity even while at rest, aptly termed resting state activation. Of recent interest is the question of how such patterns of resting state activation might predict performance on other tasks, such as lexical tone learning (Harmelech & Malach, 2013). Using the same training procedure as Chandrasekaran et al. (2011), Deng, Chandrasekaran, Wang, and Wong (2016) sought to determine if lexical tone learning success could be predicted on the basis of pre-training patterns of resting state activation. Indeed, patterns of resting state activation were predictive of lexical tone learning success, with activation localized to the left superior temporal gyrus—associated with language processing—being positively correlated with learning outcomes.

Knowing that individual variation in lexical tone learning outcomes can be traced to individual variation in neuroanatomy and neurophysiology, one question that naturally arises is where do these variations in neuroanatomy and neurophysiology come from? A population-based study demonstrated a significant correlation between the frequency of derived alleles in the *ASPM* and *MCPH1* genes and the use of lexical tone in a language, suggesting that genes may give rise to brain differences that ultimately lead to tone perception differences (Dediu & Ladd, 2007). To test this possibility, native English speakers of European descent (i.e., no ancestry in areas associated with linguistic tone) completed the phonological awareness test for pitch patterns used by Wong and Perrachione (2007), completed an fMRI procedure during which they heard repeated instances Mandarin lexical tones, and submitted a genetic sample (Wong, Chandrasekaran, & Zheng, 2012a). A derived allele from the *ASPM* gene was found to be a significant predictor of both phonological sensitivity for pitch patterns and the degree of repetition suppression to lexical tone; the predictive relationship between the derived allele and repetition suppression remained even after the listeners' phonological sensitivity for pitch patterns was factored out.

Having tied both behavioral and neurological predictors of individual variation in lexical learning outcomes to genetic variation, we close this section by returning to behavioral predictors of learning success, namely phonological sensitivity to pitch patterns. As noted by Gandour (1983), native Mandarin listeners tend to give more weight to the direction of the pitch contour whereas native English listeners tend to give more weight to pitch height. The above investigations of listeners' phonological sensitivity to pitch patterns did not differentiate whether listeners were weighting pitch height or pitch direction more heavily, though one might surmise that they were giving more weight to pitch direction, leading to success learning the lexical tone words. Prior to training, successful listeners gave more weight to pitch direction and were better able to identify pitch direction relative to less successful learners (Chandrasekaran, Sampath, & Wong, 2010), supporting the hypothesis that successful learners come to the learning task already weighting the acoustic cues within pitch patterns in a more native-like manner (see Lee & Wiener, Chap. 2, this volume, for a discussion for how non-native listeners utilize acoustic cues to perceive tone after learning). Thus, though the situation for less successful learners would initially seem bleak, as their struggles mastering lexical tone appear to have a genetic basis, having this understanding of how successful and less successful learners are differentially approaching the task prepares us to attempt to optimize the learning outcomes of the less successful learners by adjusting the training paradigms to capitalize on less successful learners' learning methods. We discuss some of these efforts in the following section.

4.3 Optimizing Learning Success

We closed the previous section by observing that, when learning to associate tone–syllable pairs with word meanings, successful learners come to the task weighting pitch direction more heavily whereas less successful learners come to the task weighting pitch height more heavily (Chandrasekaran et al., 2010). In an effort to bring less successful learners' post-training performance more in line with successful learners' post-training performance, Chandrasekaran, Yi, Smayda, and Maddox (2016) attempted to increase listeners' attention to pitch direction. When training listeners to categorize Mandarin lexical tones, listeners were told to (1) attend to pitch height, (2) attend to pitch direction, (3) attend to pitch direction and pitch height both, (4) attend to pitch direction but *not* pitch height, or (5) given no additional instructions. Those listeners who were told to attend to pitch direction showed better categorization performance post-training than did those listeners who were instructed to attend to pitch height (including height and direction), who were no different from listeners who received no instruction. The failure of listeners to receive a benefit from the pitch height instruction was attributed to native English listeners' bias toward pitch height (Gandour, 1983). Though this was a lexical tone categorization task, and not a word-learning task where the words differ on lexical tone (as in Chandrasekaran et al., 2010, 2011; Wong et al., 2007, 2008; Wong & Perrachione, 2007), it nonetheless suggests that less successful learners' performance could be improved by orienting them toward pitch direction.

In another effort to orient listeners to pitch direction, Liu et al. (2011) tested three methods of teaching listeners to identify lexical tones by directing their attention toward the contour of the pitch. Simultaneous with auditory presentation of a real word in Mandarin, spoken by a native Mandarin speaker, learners saw a visual presentation of (1) a schematic of the pitch contour and the pinyin representation of the word, (2) the number of the lexical tone and the pinyin representation of the word, or (3) a schematic of the pitch contour. Listeners were instructed to identify the tone. All participants were simultaneously enrolled in a college-level introductory Mandarin class, and participation in the experiment served as additional training in lexical tone identification. The two pinyin conditions showed faster tone learning during training, and the pinyin + pitch contour condition led to the best performance on the posttest. Because the pinyin was always presented before the auditory stimulus, the authors suggested this allowed listeners to focus exclusively on the tone and to ignore the syllable. They further suggested that schematic representations of the pitch contours led to a more robust representation of pitch direction. Again, this study did not ask listeners to use lexical tone in a lexical context, but it does provide further evidence that orienting learners to pitch dimension could lead to improved learning outcomes.

Returning to training native English listeners to associate word meaning with lexical tones (c.f. Wong & Perrachione, 2007). Perrachione, Lee, Ha, and Wong (2011) hypothesized that one potential source of difficulty for less successful listeners was the need to integrate across multiple talkers. As noted above, the high-variability phonetic training paradigm utilizes natural speech tokens produced by multiple talkers (Ingvalson et al., 2014; Iverson, Hazan, & Bannister, 2005; Logan et al., 1991). Previous work had demonstrated that multi-talker paradigms led to reduced accuracy identifying phonemes (e.g., Nygaard & Pisoni, 1998), possible due to increased cognitive cost resulting from the acoustic variability found in a multiple-talker environment (Nusbaum & Magnuson, 1997). Perrachione et al. therefore hypothesized that reducing the amount of acoustic variability across trials by using only one training talker throughout could lead to improved learning success for the less successful learners identified by Wong and Perrachione (2007). Using the phonological pitch sensitivity test developed by Wong and Perrachione, listeners were divided into either high aptitude listeners or low aptitude listeners based on their likelihood of being successful listeners on the multiple-talker paradigm. Half of each listener group was then assigned to either the multiple-talker training paradigm used previously (Wong & Perrachione, 2007) or to a similar paradigm that differed only by using one talker throughout training. All listeners were trained for eight days. The primary outcome measure was listeners' ability to identify trained words spoken by novel talkers, called the generalization test. Consistent with the studies in the previous section, low aptitude listeners assigned to the multiple-talker condition showed small gains across training sessions and poor performance on the generalization test relative to high aptitude listeners assigned to the same condition (these are the same pattern of post-training results as found in Chandrasekaran et al., 2010, 2011; Wong et al., 2007, 2008; Wong & Perrachione, 2007). However, the training condition × aptitude group interaction was significant, with high aptitude listeners assigned to multiple-talker training performing better relative to those assigned to single-talker training and, more markedly, low aptitude listeners assigned to the single-talker training performing much better than those low aptitude listeners assigned to multiple-talker training. Without explicitly focusing the low aptitude listeners on pitch direction, their performance was nonetheless improved by matching them to a condition that reduced across-trial acoustic variability. Importantly, this study highlighted the need to match both high aptitude and low aptitude listeners to their optimal training conditions, as high aptitude listeners performed best in the traditional high-variability phonetic training paradigm whereas low aptitude listeners had more learning success following single-talker training.

One point we have not yet mentioned in our review of the lexical learning training studies is that the Mandarin tones were superimposed on single syllables that were consistent with English phonotactics (Wong & Perrachione, 2007). This was done to eliminate the need for listeners to learn both unfamiliar segmental information while simultaneously learning unfamiliar suprasegmental information; by choosing segments consistent with English phonotactics, listeners could focus on learning the unfamiliar lexical tone. Additionally, though many words in Mandarin are bisyllabic, monosyllables were used to reduce the potential working memory load during training, particularly for less successful learners (e.g., Baddeley, Gathercole, & Papagno, 1998). As a result, there is the question of whether the division into successful and less successful learners will hold when listeners are tasked with bisyllabic stimuli where two different tones may be present on each syllable (or two identical tones realized as two different tones due to tone sandhi, Chang & Kuo, Chap. 7, this volume). Sadakata and McQueen (2014) tested whether Perrachione et al.'s (2011) division into high aptitude and low aptitude listeners would continue to interact with training variability when stimuli were bisyllabic and produced by native Mandarin speakers. Following five days of training on a low-, moderate-, or high-variability paradigm, they found the same training condition × aptitude group interaction as Perrachione and colleagues, in which listeners who were identified as high aptitude performed the best following high-variability paradigm whereas low aptitude listeners performed best following a low-variability paradigm. It therefore seems reasonable to conclude that the learning mechanisms and training paradigms identified here will translate beyond the pseudo-Mandarin stimuli used for laboratory investigation to real Mandarin words.

Though single-talker training led to improved learning outcomes for low aptitude listeners relative to multiple-talker training, performance on the final training day and on the generalization test was still not on par with that of high aptitude listeners. One possible means of further improving low aptitude listeners' learning performance is to bring their attention to pitch direction, as was done by Chandrasekaran et al. (2016) and Liu et al. (2011), albeit in a non-lexical context. This was the approach taken by Ingvalson, Barr, and Wong (2013). Like Perrachione et al. (2011), listeners were again divided into high aptitude listeners and low aptitude listeners. Half of each listener group completed single-talker training, chosen to optimize the learning outcomes for the low aptitude listener group. The remaining listeners were assigned to tone pre-training prior to the lexical training component. In the tone pre-training, listeners heard the same tone–syllable pairings used in the lexical training (originally developed by Wong & Perrachione, 2007) but instead of seeing a picture that represented the word's meaning, listeners saw an arrow that represented the pitch direction of the tone. Listeners were therefore trained to identify pitch direction in the same stimuli they would later learn to associate with lexical meanings. The single-talker lexical-only group was trained for eight days; the tone-training group was trained for three days on the tone-training portion and then for five days on the lexical training portion, for a total of eight days. As in Perrachione et al., the primary outcome measure was generalization to trained tokens spoken by novel talkers. Following training, the low aptitude group assigned to the tone-training condition performed significantly better on the generalization test than low aptitude listeners in the lexical-only condition whereas there was no difference in high aptitude listeners' performance regardless of condition.

In a similar study that sought to train native English listeners to learn words differing on the tones from Cantonese, Cooper and Wang (2013) also gave listeners pre-training experience identifying Cantonese tones in the context of multiple syllables produced by multiple talkers. Following tone identification training, listeners completed seven sessions of the lexical learning training from Cooper and Wang (2012). Importantly, all the listeners recruited for the tone-training study were non-musicians, allowing Cooper and Wang to compare learning performance to that of the musicians in the lexical-only condition from their earlier study. Following the lexical learning component, the non-musicians who received tone identification pre-training were able to identify words differing on Cantonese lexical tone as well as musicians who had received lexical-only training, and both groups were significantly better than non-musicians who had only received lexical training. Even within the lexical learning context, then, it appears that orienting listeners to pitch direction can substantially improve learning outcomes for less successful learners and may even mimic the effects of earlier musical training.

4.4 Future Directions for Optimizing Individual Outcomes for Lexical Tone Learning

Across all the studies discussed in the preceding sections, the participants were young (generally college-aged) listeners who reported no hearing deficits. In an increasingly globalized society, there is increasing pressure for older adults, whose hearing acuity is often lower than that of younger adults (e.g., Humes, 2013) to gain familiarity with a second language (Antoniou, Gunasekera, & Wong, 2013), raising the question as to whether older adults will have success mastering unfamiliar phonologies using the paradigms developed for younger adults, including the high-variability phonetic training paradigm. Specific to lexical tone, individuals who have congenital amusia provide an interesting test case to investigate how an inability to detect pitch changes impacts the ability to learn languages that utilize pitch phonemically. Recent work by our group has begun to investigate these areas, and we highlight some of our findings here.

Lexical tone learning by older adults. Relative to younger adults, older adults have more difficulty identifying and discriminating changes in pitch (Shen, Wright, & Souza, 2016). However, older adults also show extensive individual variation in their ability to identify and discriminate pitch patterns, not unlike the individual variation younger adults show on the phonological pitch pattern test developed by Wong and Perrachione (2007). Ingvalson, Nowicki, Zong, and Wong (2017) therefore hypothesized that older adults would show the same listener aptitude group $\times$ training condition interaction as the younger adults from Perrachione et al. (2011)

but that the older adults' performance would be attenuated overall relative to that of younger adults. One of the immediate findings was that many older adults had such difficulty identifying the pitch patterns used in the phonological pitch sensitivity test that instead of using the criterion used to divide younger adults into high aptitude listeners versus low aptitude listeners, those older adults whose performance was significantly better than chance were classified as high aptitude listeners and those whose performance was statistically equivalent to chance were classified as low aptitude listeners. As in Perrachione et al., half of each listener group was assigned to either multiple-talker training or single-talker training to learn to associate Mandarin tone–syllable pairs with lexical meanings; all listeners were trained for eight days. Unlike in younger adults, older adults' baseline phonological pitch pattern sensitivity was not a significant predictor of training performance nor of generalization performance for any listener group or any training type. Instead, baseline measures of auditory working memory and declarative memory were the best predictors of older adults' generalization performance. Comparing predictors of generalization performance between younger and older adults revealed that younger adults' generalization performance was best predicted by baseline phonological measures whereas older adults' performance was best predicted by baseline working and declarative memory measures. This suggests that not only do older adults have more difficulty perceiving pitch changes than do younger adults, but they are approaching the learning task in a fundamentally different way, in essence attempting to rote-memorize each individual stimulus item rather than extrapolate larger category features (Maddox, Chandrasekaran, Smayda, & Yi, 2013; Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010). Paradigms that lead to successful learning by older adults, then, may need to accommodate older adults' learning strategies by reducing the working memory load across training trials or induce a more efficient categorization strategy in older adult learners.

Lexical tone learning in individuals with congenital amusia. Congenital amusia affects approximately 4% of the population, including populations that use tone languages (Wong et al. 2012a, 2012b). Individuals with congenital amusia have difficulty discriminating pitch, detecting changes in pitch, identifying pitch direction, and poor memory for pitch (Ayotte, Peretz, & Hyde, 2002; Hyde & Peretz, 2004; Tillmann et al., 2010, 2011), although some of these abilities seem to be malleable (Liu, Jiang, Francart, Chan, & Wong, 2017). The reader will note that the list of tasks that individuals with congenital amusia find difficult corresponds to the list of skills found to predict lexical tone learning success in listeners who are typically developing. Given this overlap in skill set, as well as the fact that musicians show enhanced brainstem encoding of both musical and speech stimuli relative to non-musicians (e.g., Parbery-Clark, Skoe, Lam, & Kraus, 2009), Liu, Maggu, Lau, and Wong (2015) sought to investigate how native speakers of a tone language who have congenital amusia could identify lexical tone and the extent to which this identification performance corresponded to brainstem responses to musical and lexical tone stimuli. Native Cantonese speakers with congenital amusia—identified by poor performance on the Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003)—were matched to native Cantonese speakers with typical pitch perception abilities on age and amount of musical training. At the brainstem level, there was no difference in the neural response to musical stimuli nor to lexical tone stimuli between the listeners with congenital amusia and the listeners with typical pitch perception abilities. However, the listeners with congenital amusia performed much more poorly than the listeners with typical pitch perception when asked to identify lexical tone. Interestingly, in the listeners with congenital amusia, there was no correlation between the brainstem response to lexical tone and their identification performance (though the correlation did hold for listeners with typical pitch perception, c.f. Chandrasekaran et al., 2011).

A follow-up study asked native Cantonese-speaking listeners with amusia to again identify lexical tone, as well as to identify pitch direction in a speech syllable and in a piano tone then to produce a series of lexical tones and sing (Liu et al., 2016). Native Cantonese speakers with typical pitch perception also completed all tasks and served as controls. Participants' productions—both lexical tones and singing—were judged by native Cantonese listeners who were naïve to the experiment and were acoustically analyzed for pitch accuracy and pitch direction. As in the previous study, listeners with congenital amusia were less accurate than listeners with typical pitch perception at identifying lexical tone; they were also less accurate at identifying pitch direction both in speech and in piano tones. The songs by participants with congenital amusia were judged to be less accurate, and acoustic measurements demonstrated that their songs included more pitch errors. However, the lexical tone productions by the participants with congenital amusia were not less recognizable nor were they acoustically distinguishable from those produced by the participants with typical pitch perception. This pattern of results presents an interesting puzzle for investigators going forward: In native English speakers learning lexical tone, encoding at the brainstem was a predictor of future learning success (Chandrasekaran et al., 2011), but the brainstem response and tone identification are not associated for individuals with congenital amusia. Further, one of the earliest studies training listeners to identify lexical tone demonstrated that perceptual training was sufficient to improve production performance (Wang, Jongman, & Sereno, 2003a), but though individuals with congenital amusia's lexical tone identification are impaired, their production is typical. A further exploration of tone perception and tone production by individuals with congenital amusia can be found in Ong, Tan, Chan, and Wong (Chap. 8, this volume). Better understanding of where the dissociation between the brainstem's encoding and the behavioral response occurs, and the relationship between perception and production, will provide insight for improving lexical tone learning for both listeners with congenital amusia living in a tone language environment as well as native English listeners who may struggle to master a lexical tone using the training paradigms identified to date.

4.5 Conclusions

Over the course of this chapter, we have presented a series of studies demonstrating first that it is possible for native English listeners to learn lexical tone, though there may be some limitations on the ability to do so in a statistical learning context. As is often the case in non-native phoneme learning, there are extensive individual differences in learning outcomes. We traced the sources of these individual differences in learning outcomes to individual variation in neuroanatomy, neurophysiology, and genetic expression, which can be assessed via a baseline test of phonological sensitivity for pitch patterns. Additional sources of variability were traced to listeners' musical background. Recognizing that successful learners place more weight on pitch direction relative to less successful learners, who place more weight on pitch height, efforts that draw less successful learners' attention to pitch direction can optimize learning success for this group. At the same time, successful learners do the best with the traditional multiple-talker paradigm, demonstrating the importance of identifying, at baseline, which listeners are likely to be successful learners and which listeners are likely to be less successful learners in order to match learners to the correct training paradigm. We ended this chapter by presenting some new challenges for optimizing individual outcomes, older adult learners and individuals with congenital amusia, both of whom appear to approach the task of learning and perceiving lexical tone very differently from the listeners who participated in our previous work. We look forward to continuing to investigate these challenges, and encourage other researchers to recognize the importance of optimizing learning outcomes for each individual learner.

Acknowledgements The authors would like to acknowledge the support from the National Institutes of Health (USA) (R21DC016069), and the Research Grants Council (HKSAR) (34000118).

References

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians (pp. 397–400), Lisbon, Portugal.

Antoniou, M., Gunasekera, G. M., & Wong, P. C. M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience & Biobehavioral Reviews*, *37*(10, Part 2), 2689–2698. https://doi.org/10.1016/j.neubiorev.2013.09.004.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain: A Journal of Neurology*, *125*(Pt 2), 238–251.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English/r/ and/l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivity to lexical tone in an artificial language. *Studies in Second Language Acquisition, 37*(Special Issue 02), 335–357. https://doi.org/10.1017/S0272263114000849.

Chandrasekaran, B., Kraus, N., & Wong, P. C. M. (2011). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology, 107*(5), 1325–1336. https://doi.org/10.1152/jn.00923.2011.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456–465. https://doi.org/10.1121/1.3445785.

Chandrasekaran, B., Yi, H.-G., Smayda, K. E., & Maddox, W. T. (2016). Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics, 78*(2), 566–582. https://doi.org/10.3758/s13414-015-0999-x.

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America, 139*(5), 2432–2447. https://doi.org/10.1121/1.4947497.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America, 131*(6), 4756–4769. https://doi.org/10.1121/1.4714355.

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *The Journal of the Acoustical Society of America, 134*(2), EL133–EL139. https://doi.org/10.1121/1.4812435.

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences, 104*(26), 10944–10949. https://doi.org/10.1073/pnas.0610848104.

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. M. (2016). Resting-state low-frequency fluctuations reflect individual differences in spoken language learning. *Cortex, 76,* 63–78. https://doi.org/10.1016/j.cortex.2015.11.020.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11,* 149–175.

Harmelech, T., & Malach, R. (2013). Neurocognitive biases and the patterns of spontaneous correlations in the human cortex. *Trends in Cognitive Sciences, 17*(12), 606–615. https://doi.org/10.1016/j.tics.2013.09.014.

Humes, L. E. (2013). Understanding the speech-understanding problems of older adults. *American Journal of Audiology, 22,* 303–305. https://doi.org/10.1044/1059-0889(2013/12-0066).

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356–360. https://doi.org/10.1111/j.0956-7976.2004.00683.x.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research, 56*(3), 1045–1050. https://doi.org/10.1044/1092-4388(2012/12-0024).

Ingvalson, E. M., Ettlinger, M., & Wong, P. C. M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism, 18*(1), 35–47. https://doi.org/10.1177/1367006912456586.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r–l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*(2), 255–274. https://doi.org/10.1017/S1366728911000447.

Ingvalson, E. M., Nowicki, C., Zong, A., & Wong, P. C. M. (2017). Non-native speech learning in older adults. *Frontiers in Psychology, 8.* https://doi.org/10.3389/fpsyg.2017.00148.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866. https://doi.org/10.1121/1.3148196.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*(5), 3267–3278. https://doi.org/10.1121/1.2062307.

Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*(2), 349–366. https://doi.org/10.1111/bjop.12092.

Liu, F., Chan, A. H., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. (2016). Pitch perception and production in congenital amusia: Evidence from Cantonese speakers. *The Journal of the Acoustical Society of America, 140*(1), 563–575.

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., & Wong, P. C. M. (2017). Perceptual Learning of pitch direction in congenital amusia: Evidence from Chinese speakers. *Music Perception: An Interdisciplinary Journal, 34*(3), 335–351. https://doi.org/10.1525/mp.2017.34.3.335.

Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. *Frontiers in Human Neuroscience, 8,* 1029. https://doi.org/10.3389/fnhum.2014.01029.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning, 61*(4), 1119–1141. https://doi.org/10.1111/j.1467-9922.2011.00673.x.

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*(3), 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874–886.

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging, 28*(4), 1042–1056. https://doi.org/10.1037/a0034969.

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia, 48*(10), 2998–3008. https://doi.org/10.1016/j.neuropsychologia.2010.06.008.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. *Talker Variability in Speech Processing*, 109–132.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376.

Ong, J. H., Burnham, D., & Escudero, P. (2015). Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PloS One*, *10*(7), e0133446.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear & Hearing, 30*(6), 653–661.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences, 999,* 58–75.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America, 130*(1), 461–472. https://doi.org/10.1121/1.3593366.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Language Sciences, 5,* 1318. https://doi.org/10.3389/fpsyg.2014.01318.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Sebastian-Galles, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72,* 111–123.

Shen, J., Wright, R., & Souza, P. E. (2016). On older listeners' ability to perceive dynamic pitch. *Journal of Speech Language and Hearing Research, 59*(3), 572. https://doi.org/10.1044/2015_JSLHR-H-15-0228.

Shen, X. (1989). Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics, 17*(1), 61–74.

Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The amusic brain: Lost in music, but not in space. *PLoS ONE, 5*(4). https://doi.org/10.1371/journal.pone.0010173.

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., & Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *The Journal of the Acoustical Society of America, 130*(6), 4089–4096. https://doi.org/10.1121/1.3658447.

Wang, Y., Jongman, A., & Sereno, J. A. (2003a). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America, 113*(2), 1033. https://doi.org/10.1121/1.1531176.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003b). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience, 15*(7), 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*(6), 3649–3658.

Werker, J. F., & Stager, C. L. (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In M. Broe & J. Pierre-Humbert (Eds.), *Papers in laboratory phonology* (Vol. 5, pp. 181–193). Cambridge: Cambridge University Press.

Wong, P. C. M., Chandrasekaran, B., & Zheng, J. (2012a). The derived allele of ASPM is associated with lexical tone perception. *PLoS ONE, 7*(4), e34243. https://doi.org/10.1371/journal.pone.0034243.

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L. H., & Peretz, I. (2012b). Effects of culture on musical pitch perception. *PLoS ONE, 7*(4), e33424.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*(04), 565–585. https://doi.org/10.1017/s0142716407070312.

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping, 28*(10), 995–1006. https://doi.org/10.1002/hbm.20330.

Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex, 18*(4), 828–836. https://doi.org/10.1093/cercor/bhm115.

Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., & Shu, H. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Auditory Cognitive Neuroscience*, 436. https://doi.org/10.3389/fpsyg.2015.00436.