Published in final edited form as:

J Speech Lang Hear Res. 2013 June ; 56(3): 1045–1050. doi:10.1044/1092-4388(2012/12-0024).

Poorer Phonetic Perceivers Show Greater Benefit in Phonetic-Phonological Speech Learning

Erin M. Ingvalson^a, Allison M. Barr^a, and Patrick C. M. Wong^{a,b}

^aNorthwestern University, Evanston, IL

^bThe Chinese University of Hong Kong, SAR, The People's Republic of China

Abstract

Purpose: Previous research has demonstrated that native English speakers can learn lexical tones in word context (pitch-to-word learning), to an extent. However, learning success depends on learners' pre-training sensitivity to pitch patterns. The aim of this study was to determine whether lexical pitch-pattern training given before lexical training could improve learning and whether or not the extent of improvement depends on pre-training pitch-pattern sensitivity.

Method: Learners with high and low pitch-pattern sensitivity were given training on lexical pitch patterns before lexical training.

Results: It was found that such training resulted in better learning than lexical training alone, primarily in learners with low pre-training pitch-pattern sensitivity.

Conclusion: These data support the importance of considering individual aptitudes when developing training and also the notion of phonetic–phonological–lexical continuity in word learning.

Keywords

phonology; language; speech perception

For adults, learning a new language is both time-consuming and challenging. Recent work (e.g., Golestani & Zatorre, 2009; Perrachione, Lee, Ha, & Wong, 2011; Wong & Perrachione, 2007) has demonstrated that some learners may face a greater challenge due to a low perceptual aptitude for some nonnative contrasts. Recent work has also demonstrated that these individual differences in ability can be overcome, to an extent, through the development of training paradigms that take into account a learner's initial aptitude (Perrachione et al., 2011).

Native English speakers, for whom pitch patterns do not convey lexical information, can be successfully trained to learn the lexical tones of Mandarin Chinese (Wang, Spence, Jongman, & Sereno, 1999). There is suggestion that a phonological awareness of tone is necessary for successful word learning for both native English (Wong & Perrachione, 2007) and native Mandarin speakers (So & Siegel, 1997). Poor phonological awareness also

impedes new word learning in both the first (Ho, Chan, Tsang, Lee, & Chung, 2006) and second (Hu, 2003; Hu & Catts, 1998) languages. Training on phonological awareness can improve lexical learning in a broad group of learners (Cooper & Wang, 2011). Wong and Perrachione (2007) demonstrated that learners' initial sensitivity to pitch patterns (in a nonlexical context) successfully predicted ultimate attainment when learning a nonnative phonemic (tonal) distinction in lexical context. It therefore follows that making learners with poor phonological awareness more aware of the crucial sounds—in this case, lexical tone—of the to-be-learned words should lead to greater learning success.

In this study, learners with high and low initial sensitivity to pitch patterns participated in a training study in which they first received lexical pitch-pattern training (for 3 days) followed by lexical pitch-to-word learning (for 5 days) across a total of 8 days. Their results were compared with those of learners with high and low initial sensitivity to pitch patterns who participated in an 8-day lexical pitch-to-word learning task (Perrachione et al., 2011). We hypothesized that focusing the listeners on the tone contrasts themselves would benefit those learners with a low aptitude for tone learning by attuning them to the crucial distinction.

Materials and Method

Participants

Forty-one normal-hearing, native English-speaking participants ages 18 to 21 were recruited from the Northwestern University community. None had previous experience with a tonal language. Twenty-one participants were assigned to the pitch-sensitivity condition, described below. The remaining 20 participants were assigned to lexical-only training and were those participants who completed the blocked-by-talker training in Perrachione et al. (2011).

Materials

Wong and Perrachione (2007; see also Perrachione et al., 2011) recorded a set of six syllables (*dree*, *nuck*, *fute*, *pesh*, *nare*, and *vess*) each produced once by eight native American English speakers (four female, four male). These syllables were then superimposed with three pitch contours (level, rising, and falling) using the Pitch Synchronous Overlap and Add (PSOLA) method in Praat. This resulted in 18 pseudowords (six syllables $\times$ three pitch contours; see Table 1) per talker. These 18 pseudowords were associated with 18 common objects (e.g., cow, phone, bus), for the word-learning portion of the experiment. Four talkers (two female, two male) were designated for use in training; the remaining four talkers (two female, two male) were designated for use in a generalization test (called the Test of Learning Achievement [TLA; Perrachione et al., 2011]).

Resynthesis was applied across the sonorous segments of each token, using each talker's mean F0 across all productions as a baseline value. The level contour began and ended at the baseline pitch; the rising contour began at 74% of baseline and ended at the baseline pitch; the falling contour began at 110% of baseline and fell by 45%. The resulting contours are in line with measurements obtained by Shih (1988), and have been shown to be perceptually natural to native Mandarin speakers (Wong & Perrachione, 2007). Onset and offset values

were varied by $\pm 3\%$ across talkers to produce talker-specific variability in the target phonemic contrast.

Procedure

Prior to training, all participants completed the Pitch Contour Perception Test (PCPT; Wong & Perrachione, 2007). This test has been found to be highly predictive of listeners' ability to learn lexical tone (Perrachione et al., 2011; Wong & Perrachione, 2007). In the test, listeners hear 180 tokens of isolated vowels, each superimposed with a level, rising, or falling pitch contour (as in the training and testing stimuli, resynthesis was done using PSOLA in Praat; Wong & Perrachione, 2007). Contours on these vowels followed the same parameters as on the full-syllable stimuli. Listeners identified the direction of the contour by matching the auditory stimulus with representative arrows presented on a computer monitor ($\rightarrow$, $\nearrow$, or $\searrow$) in a two-alternative forced-choice task (only two of the three possible response options were presented on any given trial). Listeners who scored greater than 70% were considered high-aptitude learners (HALs); listeners who scored at or below 70% were considered low-aptitude learners (LALs). Previous work has indicated that this division is highly reliable in predicting which listeners will master the lexical tone contrast following training versus which listeners will show a plateau in training performance at below-optimum levels (Perrachione et al., 2011; Wong & Perrachione, 2007). Ten listeners in this study were classified as HALs and 11 listeners were classified as LALs; there were 10 listeners of each type in the lexical-only group.

All training was done using the pseudoword stimuli described above. The only difference between the tone-training component and the lexical component was the listeners' task. The first 3 days of training comprised the pitch-sensitivity component. Pitch-sensitivity training was done within syllable to emphasize the difference among the pitch contrasts, and listeners received feedback on their response accuracy. During this segment of training, listeners were trained to identify the direction of the pitch contour in each syllable. Over a triad of trials, listeners would hear three instances of the same syllable spoken by the same talker with three different pitch contours. An arrow depicting the direction of the contour would appear in conjunction with the syllable. Once all three pitch contours had been presented, listeners were asked to identify the pitch contour of a syllable (three-alternative forced choice), and feedback on their accuracy was given. Listeners, therefore, learned stimuli as minimal triplets, although stimuli were presented in isolation. Listeners heard all tokens from all talkers in each training session (72 total tokens), and tokens were blocked by talker. Each day's training session ended with an identification test that asked listeners to identify a stimulus token's pitch pattern, with all tokens from all talkers randomly intermixed. No feedback was given during each day's final identification test.

The remaining 5 days of training comprised the lexical only component. The procedure for this segment of training was very similar to the pitch-sensitivity component, except that the visual stimuli were the pictures that depicted the lexical meanings of the pseudowords (e.g., *cow*, *phone*, *bus*), which emphasized the relationship between the pitch patterns and their lexical meanings. All other aspects of training were identical to the pitch-sensitivity component. In each day's final identification test, listeners heard a single stimulus token and

were asked to identify its meaning from the entire set of 18 possible meanings (i.e., 18-alternative forced choice). A visual display with all 18 possible meanings was continuously present, and listeners chose their responses from this display following each stimulus presentation. Only lexical meanings were shown in the display; no other information was presented. All tokens from all talkers were presented in random order, and no feedback was given on the final identification test. Lexical-only trained listeners received training on the lexical training procedure for all 8 days (see Perrachione et al., 2011, Experiment 2).

After completing the 8 days of training, both the pitch-sensitivity and lexical-only groups returned on a separate day for the TLA (Perrachione et al., 2011). This test was very similar in structure to the final identification test used during the lexical training component, except that the four talkers reserved for generalization were used. As before, listeners heard one pseudoword—this time always spoken by an unfamiliar talker—and identified its lexical meaning from the complete list of 18 items. Each stimulus was presented only once, and no feedback was given.

Results

Before analysis, all data were arcsine transformed (Studebaker, 1985). Because the pitch-sensitivity group and lexical-only groups were recruited separately, we first needed to verify that their subdivisions into HAL and LAL were similar. Prior to training, the HAL groups for pitch-sensitivity and the lexical-only trained listeners were not significantly different on their PCPT performance, $t(18) = 0.10$, $p = .92$ (proportion correct pitch sensitivity, $M = 0.91$; lexical only, $M = 0.91$). The same was true for the pitch-sensitivity and lexical-only LAL groups before training on the PCPT, $t(19) = 0.48$, $p = .64$ (proportion correct pitch sensitivity, $M = 0.58$; lexical only, $M = 0.59$). We, therefore, proceeded with the analyses.

The learning curves are shown in Figure 1. HAL were successful under both conditions, but the pitch-sensitivity group appeared to make more rapid progress and asymptote more quickly on the lexical component than the lexical-only group. Entering the first 3 lexical training days from each group into an analysis of variance (ANOVA) for linear trend indicated that the training progression for each group is significant—pitch sensitivity, $F(1, 63) = 53.10$, $p < .001$, $\eta_p^2 = .18$; lexical only, $F(1, 63) = 157.10$, $p < .001$, $\eta_p^2 = .81$ —but the lexical-only group's curve does appear to be steeper, suggesting less rapid progress. There is also some suggestion that the pitch-sensitivity HAL group members were not notably different from final attainment on their second day of lexical training (Training Day 5 vs. Training Day 8), $t(9) = 2.12$, $p = .06$, although further work is needed to determine whether training could be shortened for this group without diminishing training gains. Although the pitch-sensitivity LAL group did not begin lexical training until Training Day 4, upon visual inspection, it appears that they quickly surpassed the lexical-only group—who showed a leveling off of performance around Training Day 5—and that gains continued throughout the course of lexical training.

A 2×2 (Group $\times$ Condition) ANOVA was conducted on TLA scores to determine final learning performance (see Figure 2). There was a main effect of group, $F(1, 37) = 41.12$, p

$< .001$, $\eta_p^2 = .57$. The HAL group performed better than the LAL group in both the tone-training, $F(1, 19) = 23.69$, $p < .001$, $\eta_p^2 = .55$, and lexical-only, $F(1, 18) = 21.74$, $p < .001$, $\eta_p^2 = .55$, conditions. There was also a main effect of condition, $F(1, 37) = 4.95$, $p = .03$, $\eta_p^2 = .12$, indicating that pitch-sensitivity training resulted in better generalization performance than lexical-only training. We found the expected interaction of group and training type, $F(3, 37) = 16.44$, $p < .001$, $\eta_p^2 = .08$. Simple effects tests revealed that the pitch-sensitivity trained LAL group outperformed the lexical-only LAL group, $F(1, 19) = 4.84$, $p = .04$, $\eta_p^2 = .20$, whereas there was no effect of training condition on the HAL group, $p = .62$. The performance of both HAL groups was high enough to cause concern of a ceiling effect. Comparing the performance of the two HAL groups to arcsine-transformed perfect identification revealed that both groups were significantly different from perfect—pitch sensitivity, $t(9) = 5.13$, $p < .001$, $d = 0.86$; lexical only, $t(9) = 6.07$, $p < .001$, $d = 0.90$ —indicating that neither group was at absolute accuracy.

A $2 \times 2 \times 2$ (Group $\times$ Condition $\times$ Session) mixed-model ANOVA conducted on the PCPT scores to determine the effect of pitch-sensitivity training on tone perception (see Figure 3) revealed a significant main effect of group, $F(1, 37) = 50.40$, $p < .001$, $\eta_p^2 = .44$, indicating that the HAL group showed better tone perception regardless of training condition or test session. There was also a main effect of session, $F(1, 37) = 5.89$, $p = .02$, $\eta_p^2 = .06$, where posttest performance overall was better than pretest performance. The three-way interaction was also significant, $F(1, 37) = 4.75$, $p = .04$, $\eta_p^2 = .05$. The simple effects tests revealed that the effect of session was significant only for the pitch-sensitivity trained LAL group, $F(1, 20) = 49.86$, $p < .001$, $\eta_p^2 = .71$, where perception for the LAL group improved from pretest (proportion correct $M = 0.58$) to posttest (proportion correct $M = 0.83$), and no other groups showed a significant pre- to posttest change. Interestingly, post hoc Tukey tests revealed no significant differences between the pitch-sensitivity trained LAL group and either HAL group on PCPT scores at posttest, whereas the lexical-only trained LAL group continued to perform significantly lower than the HAL groups.

Thus, training type did not appear to impact the generalization performance of the HAL group, whose pre-training tone sensitivity was maintained throughout training; this pre-training tone sensitivity allowed for successful learning regardless of the method used. For this group, however, training type may impact the rate of training progress, with pitch-sensitivity trained listeners beginning lexical training at a higher accuracy rate and reaching asymptote more quickly than lexical-only trained listeners. The significant main effect of training condition on generalization scores additionally suggests that the HAL learners may receive some benefit from pitch-sensitivity training, although outcomes will be equivalent regardless of training condition. Conversely, for the LAL group, the type of training mattered a great deal for posttest performance. Those listeners who received pitch-sensitivity training showed a significant improvement on their tone-perception ability, as measured by the PCPT. Those listeners who received pitch-sensitivity training also showed greater lexical

learning than did their lexical-only trained counterparts, presumably as a result of their improved tone-perception sensitivities (Wong & Perrachione, 2007).

Discussion

These findings replicated and extended on the previous findings of Wong and Perrachione (2007) and of Perrachione et al. (2011) that learners' ultimate lexical tone-learning success can be successfully predicted by their pre-training sensitivity to pitch patterns and that not all training paradigms benefit all learners equally. They also expand on the findings of Cooper and Wang (2011), who found a benefit of phonological pre-training for all learners but did not examine the particular benefits to LALs. The HAL group successfully learned the lexical tones in both the pitch-sensitivity and lexical-only conditions (see also Perrachione et al., 2011, for more conditions under which HAL learners are successful). Conversely, the LAL group showed a greater benefit when the initial portion of training emphasized tone identification before moving on to lexical learning than when training focused on lexical learning alone. The lexical learning condition of our lexical-only trained participants—blocked-by-talker presentation—had been previously found to be the most beneficial condition for LAL (Perrachione et al., 2011), making it notable that initially focusing on learners' lexical pitch-pattern sensitivity appears to provide this learner group with an even greater benefit. Because pitch-sensitivity training also improved the LAL group's pitch-pattern perception, it is likely that the additional benefit is a direct result of improved sensitivity to pitch patterns. It is also worth noting that the pitch-sensitivity trained listeners' performance did not appear to asymptote after 5 days of lexical training, leaving open the possibility that if pitch-sensitivity training was followed by 8 days of lexical-only training—the amount received by the lexical-only training groups—then the tone-trained LAL listeners may have continued to improve.

There is evidence indicating that infants have difficulty using phonemic minimal pair distinctions when learning new words, although they are capable of such phonetic discriminations in phonological tasks (e.g., Stager & Werker, 1997). This evidence has been used to suggest that in infant word learning, the level of phonetic detail used for word learning is not the same as that used for simple speech perception tasks (Werker & Stager, 2000). From the perspective of lexical learning in infancy, this strategy could be advantageous, as many words are distinguished by multiple phones, and it is only as the lexicon grows that greater attention to finer phonetic detail is needed to distinguish among minimal pairs (Werker & Stager, 2000), suggesting that greater phonetic detail would be used for word learning at older ages. It is possible that weaker adult language learners show a pattern similar to that seen in the infants and also do not differentiate words, including new to-be-learned words, on the basis of fine phonetic detail, although they are capable of making the phonological distinction in a nonlexical context (e.g., Wong & Perrachione, 2007), which, in turn, hinders their ability to acquire the new lexicon. This would suggest that improving such learners' ability to differentiate phonetic detail in a lexical context would improve their ability to acquire a new lexicon. Although the study here focused on pitch-pattern sensitivity, previous evidence suggests that pitch-pattern sensitivity is related to tone awareness, which is phonologically contrastive for many languages (e.g., Wong & Perrachione, 2007). This study, therefore, suggests the possibility that heightening weaker

adult language listeners' awareness for the relevant phonological distinction can improve lexical learning.

It is possible that the learning here is not a result of explicit instruction on the to-be-learned phonological distinction but is instead a result of passive exposure to the lexical stimuli without an explicit lexical task followed by explicit lexical learning (so-called "latent learning"; see, e.g., Thistlethwaite, 1951). Based on earlier work demonstrating the importance of phonological awareness for word learning (e.g., Ho et al., 2006; C. Hu, 2003; C.-F. Hu & Catts, 1998), we believe that the initial training is highlighting the phonological distinction. As we work toward our goal of developing personalized training paradigms, we recognize that we will need to definitively determine the extent to which learning gains are influenced by passive exposure to lexical stimuli versus explicit phonological training, and we set this as a goal for future research.

Conclusions

This study demonstrated that pitch-pattern sensitivity training can improve lexical learning. Learners with both high and low initial sensitivity for the to-be-learned contrast showed a benefit of training, marked by more rapid progress during training by HALs and by higher overall outcomes by LALs. Although their initial aptitudes appear to prepare the HALs for good outcomes regardless of the training type, these data suggest the possibility that training time could be shortened for these high-ability learners by an initial focus on pitch patterns. For the LALs, the benefits of pitch-pattern sensitivity training are more pronounced, bringing this group of learners' pitch identifications in line with those of HALs and improving their overall lexical learning. We believe that the benefit that LALs show from initial pitch-pattern training can generalize to phonological awareness training overall and highlights the importance of phonological awareness for lexical learning, particularly the importance of phonological awareness for the acoustic cue that differentiates the to-be-learned items. These data can also be added to the growing body of evidence that learners' initial abilities should be taken into account when designing secondlanguage learning paradigms. We welcome other researchers to join us in the development of such personalized paradigms.

Acknowledgment

This research was funded by National Institutes of Health Grants R01DC008333 and K02AG03582 and National Science Foundation Grant BCS-1125144, all of which were awarded to the third author.

References

- Cooper A, & Wang Y (2011). The influence of tonal awareness and musical experience on tone word learning. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 512–515). Hong Kong, SAR, China: ICPHS.
- Golestani N, & Zatorre RJ (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, 109, 55–67. [PubMed: 18295875]
- Ho CS-H, Chan DW, Tsang S-M, Lee S-H, & Chung KKH (2006). Word learning deficit among Chinese dyslexic children. *Journal of Child Language*, 33, 145–161. [PubMed: 16566324]
- Hu C (2003). Phonological memory, phonological awareness, and foreign language word learning. *Language Learning*, 53, 429–462.

- Hu C-F, & Catts HW (1998). The role of phonological processing in early reading ability: What we can learn from Chinese. *Scientific Studies of Reading*, 2, 55–79.
- Perrachione TK, Lee J, Ha LYY, & Wong PCM (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130, 461–472. [PubMed: 21786912]
- Shih C (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.
- So D, & Siegel LS (1997). Learning to read Chinese: Semantic, syntactic, phonological and working memory skills in normally achieving and poor Chinese readers. *Reading and Writing*, 9, 1–21.
- Stager CL, & Werker JF (1997, 7 24). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381–382. [PubMed: 9237755]
- Studebaker GA (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, 28, 455–462. [PubMed: 4046587]
- Thistlethwaite D (1951). A critical review of latent learning and related experiments. *Psychological Bulletin*, 48, 97–129. [PubMed: 14834282]
- Wang Y, Spence MM, Jongman A, & Sereno JA (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106, 3649–3658. [PubMed: 10615703]
- Werker JF, & Stager CL (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In Broe M & Pierre-Humbert J (Eds.), *Papers in Laboratory Phonology 5* (pp. 181–193). Cambridge, England: Cambridge University Press.
- Wong PCM, & Perrachione TK (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28, 565–585.

Figure 1.

Arcsine-transformed accuracy on each training day's final test for the high-aptitude learners (left panel) and low-aptitude learners (right panel) as a function of training type, either starting with initial tone learning (pitch-sensitivity group) or focusing only on lexical learning (lexical-only group). The break in the pitch-sensitivity group's curves indicates the shift from phonological training to lexical training. Error bars represent standard error of the mean (SEM).

Figure 2.

Arcsine-transformed accuracy on the Test of Learning Achievement (Perrachione et al., 2011) as a function of learning aptitude (high-aptitude learner or low-aptitude learner) and training type (pitch sensitivity, focusing on sensitivity to lexical pitch patterns then lexical training, or lexical only). The dashed line indicates perfect identification performance. Error bars represent *SEM*.

Figure 3. Arcsine-transformed accuracy on the Pitch Contour Perception Test (Wong & Perrachione, 2007) as a function of learning aptitude, training type, and test session. Only the low-aptitude learners who received pitch-sensitivity training showed a significant change from pre- to posttest. Error bars represent *SEM*.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Artificial vocabulary trained in the lexical component.

p ^h ɛŋ1 (glass)	d.li11 (arm)	ne.r1 (boat)	ves1 (hat)	n.ʌk1 (brush)	fjʉt1 (shoe)
p ^h ɛŋ2 (pencil)	d.li12 (phone)	ne.r2 (potato)	ves2 (tape)	n.ʌk2 (tissue)	fjʉt2 (book)
p ^h ɛŋ4 (table)	d.li14 (cow)	ne.r4 (dog)	ves4 (piano)	n.ʌk4 (bus)	fjʉt4 (knife)

Note. Numbers following the lexical items designate tone. Level tone is indicated by 1, rising tone by 2, and falling tone by 4. During lexical training, word meanings were represented pictorially.