

Changes in early L2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese

Seth Wiener

Carnegie Mellon University, Pittsburgh, PA, USA

sethw1@cmu.edu

Abstract

This study examined how cue-weighting of a non-native speech cue changes during early adult second language (L2) acquisition. Ten native English speaking learners of Mandarin Chinese performed a speeded AX-discrimination task during months 1, 2, and 3 of a first-year Chinese course. Results were compared to ten native Mandarin speakers. Learners' reaction time and d-prime results became more native-like after two months of classroom study but plateaued thereafter. Multidimensional scaling results showed a similar shift to more native-like cue-weighting as learners attended more to pitch direction and less to pitch height. Despite the improvements, learners' month 3 configuration of cue-weighting differed from that of native speakers; learners appeared to weight pitch end points rather than overall pitch directions. These results suggest that learners' warping of the weights of dimensions underlying the perceptual space changes rapidly during early acquisition and can plateau like other measures of L2 acquisition. Previous perceptual learning studies may have only captured initial L2 perception gains, not the learning plateau that often follows. New methods of perceptual learning, especially for tonal languages, are needed to advance learners off the plateau.

Index Terms: cue-weighting, Mandarin Chinese, second language acquisition, lexical tone

1. Introduction

Languages differ in the acoustic cues used to convey segmental (e.g., consonants and vowels) and suprasegmental (e.g., pitch) information. For adult second language (L2) learners, perceiving the cues necessary for non-native speech sound categorization is notoriously difficult; learners must change or warp how they weight acoustic cues [1]. This change may require an increase in emphasis on relevant dimensions and/or a decrease in emphasis on irrelevant dimensions [2]. Multidimensional scaling (MDS) approaches allow for an examination of the cues and dimensions that listeners use to create an underlying perceptual space [3,4]. Importantly, MDS studies have demonstrated that cue-weighting is language specific and sensitive to individual differences [4,5].

This study makes use of multidimensional scaling to examine the early development of cue-weighting in an L2. In particular, this study investigates how non-native learners of Mandarin Chinese shift their weighting of cues necessary for lexical tone categorization over the course of the first three months of classroom-based L2 learning. In Mandarin, unlike English and other European languages, a speaker's pitch variations convey phonologically contrastive information. Listeners must categorize voice fundamental frequency (F0)

information into one of four tone categories: high level (T1), high rising (T2), low dipping (T3), and high falling (T4). Native Mandarin listeners accomplish this by attending to a dimension related to pitch direction [3]. In contrast, speakers of a non-tonal first language, such as English, struggle to accurately categorize L2 tones because they attend more to a dimension related to pitch height [4]. This difference between native and non-native speakers reflects the role of pitch in each language, i.e., its lexical role in Mandarin and its non-lexical (though speaker-specific) role in English [2]. Previous studies have established that short perceptual training can improve naïve listeners' tone categorization by warping the perceptual space [1,6]. Moreover, listeners who place a stronger emphasis on pitch direction dimensions tend to be better Mandarin learners than those who place a weaker emphasis on pitch direction and a stronger emphasis on pitch height [4,7,8].

The present study advances previous research on L2 lexical tone cue-weighting in two ways. First, this study tests a motivated group of classroom learners engaged in a daily, university-level Elementary Chinese language course. L2 classroom learners bring with them a different learning attitude and motivation than naïve participants traditionally used in laboratory-based learning studies [9,10]. For instance, [4] taught naïve listeners Mandarin-like sound-image pairs over nine sessions. Results indicated a high degree of variability in individual word identification performance. Given the large individual difference, the authors separated the data in order to analyze "good" and "poor" learning groups. Classroom learners' greater motivation and increased exposure to L2 speech, should result in less overall variability during testing. This will allow for a more uniform pattern (i.e., a larger dataset of what [4] calls "good" learners) and serve as a natural litmus test of previous MDS cue-weighting findings, e.g., [3,4,6].

Second, this study examines the trajectory of learning across the first three months of L2 acquisition. Unlike previous studies that have focused on short, input-intensive training sessions across one or two weeks, e.g., [4,7], the present study tracks learners' development during a typical classroom-based learning environment as is the norm in adult L2 acquisition. This allows for a more natural development of L2 learners' change in cue-weighting. Additionally, the inclusion of a third month serves as a post-test of the cue-weighting changes reported in [4]. L2 acquisition rarely exhibits a smooth transition from beginner to intermediate to advanced levels of proficiency; initial gains are often followed by learning plateaus [11,12]. It remains an open question whether L2 learners' increased emphasis on pitch direction continues to become more native-like or whether it plateaus similar to that of other measures of L2 acquisition, e.g., [11,12].

2. Experiment

2.1. Methods

2.1.1. Participants

Ten native American English speakers (4 females, 6 males; mean age: 20.3) studying Chinese as a second language at Carnegie Mellon University participated in the study. All participants were undergraduate students with normal hearing, no prior experience with a tonal language, and no formal musical training.

Ten native Mandarin speakers (5 females, 5 males; mean age 26.2) from mainland China served as a control group. All participants were undergraduate or graduate students with normal hearing.

2.1.2. Stimulus creation

The four F0 contours were modeled on natural citation-form Mandarin F0 contours of two native speakers of Mandarin from China (one male, one female). Citation-form tones of the Mandarin vowel /y/ were recorded in a sound-attenuated chamber and sampled at 44.1 kHz. Following [4,13], the citation-form contours were separately superimposed on the vowel /y/ using the pitch synchronous overlap and add (PSOLA) method in the software Praat [14]. This resulted in four Mandarin syllable-tone combinations spoken by a male and female: *yu1* (Chinese surname Yu), *yu2* ('fish'), *yu3* ('rain'), *yu4* ('jade'). The syllable *yu* was chosen because students had learned the morphemes for these four syllable-tone combinations in their textbook. All stimuli had a normalized duration of 400 ms.

2.1.3. Procedure

Participants were tested individually in a quiet space using headphones. The task consisted of speeded AX-discrimination judgments of Mandarin tone pairs. Using a keyboard, participants indicated whether the tone pairs presented were "same" or "different." Participants were first presented with a practice set of trials. Each trial consisted of a pair of stimuli spoken by the same speaker and a 500 ms interstimulus interval. Participants were told to respond as quickly and as accurately as possible with a 1 second timeout period. Reaction time was measured from the offset of the second sound within a pair. Stimuli were presented using E-Prime with an equal probability of same and different trials and male and female trials. Different trials counterbalanced the order of the two stimuli. Participants completed 192 trials across two blocks. Correct discrimination was recorded as a 'hit' while incorrect discrimination was recorded as a 'false alarm.' If participants did not respond within 1 s, it was recorded as a 'miss.' Sensitivity index (d-prime) was calculated for each participant.

Learners underwent three testing sessions. The first session occurred during the second week of the semester in month 1. Participants were therefore still relatively unfamiliar with categorizing tone and could be considered naïve listeners similar to those participants tested in [3,4]. The second and third testing session occurred in months 2 and 3, respectively, i.e., after approximately 60 and 90 classroom hours. The native speaker control group was only tested once.

2.2. Results

2.2.1. AX-discrimination task

Figure 1 shows the mean d-prime values for native speakers and learners across the three testing sessions (error bars indicate 95% confidence intervals). Learners demonstrated an improvement from month 1 to month 2, but no change was observed from month 2 to month 3. As expected, native speakers' mean d-prime values exceeded those of learners, even after three months of classroom L2 exposure. A comparison between learners' month 3 d-prime and native speakers' d-prime values confirmed this difference was statistically significant ($t(18) = -3.43, p < .01$). To test whether learners' sensitivity significantly changed across the three months, a 1-way ANOVA was conducted. Results revealed no significant effect of month ($F(2,28) = 1.5, p > .05$).

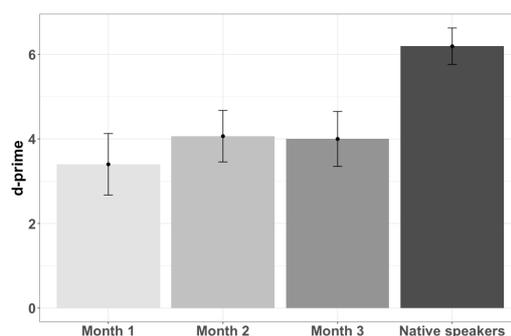


Figure 1: Mean d-prime values.

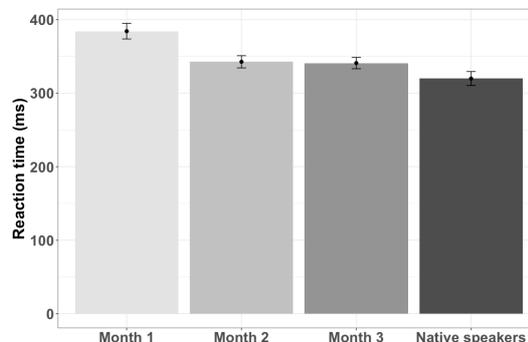


Figure 2: Mean correct reaction times.

Figure 2 shows the mean reaction times for correct responses only (with 95% confidence intervals). Overall, native speakers responded fastest while learners at month 1 responded slowest. Learners demonstrated an improvement from month 1 to month 2, but no change was observed from month 2 to month 3. Learners' reaction times at month 3 were still slower than those of native speakers. To test whether learners' reaction time changed across the three months, a mixed effects regression model was built using the *lme4* package [15] and *pbkrtest* package [16] in R (version 3.3.2). Month 2 was coded as the reference level allowing for two planned comparisons: months 1-2 and months 2-3. The random effects structure included by-subject and by-item random intercepts and slopes for month. Results revealed that

reaction times in month 2 were significantly faster than those in month 1 ($\beta = 63.476$, $SE = 18.20$, $t = 3.48$, $p < .05$). No difference was observed between reaction times in months 2 and 3 ($\beta = -0.697$, $SE = 44.46$, $t = -0.02$, $p > .05$). A separate model was built to compare month 3 reaction times to those of native speakers. Native speakers responded marginally faster than learners at month 3 ($\beta = -60.352$, $SE = 45.26$, $t = -1.99$, $p = .09$).

2.2.2. MDS analyses of reaction time data

Figure 3 plots the acoustics of the stimuli used in the discrimination task. The x-axis shows the average height in semitones for each speaker. The y-axis shows the overall slope or direction in semitones for each speaker; a positive value indicates an overall F0 rise or increase, while a negative value indicates an overall F0 fall or decrease. This plot represents the perceptual space with pitch height as the predicted first dimension (i.e., T1 and T3 can be characterized by high and low pitch heights) and pitch direction as the predicted second dimension (i.e., T2 and T4 can be characterized by rising and falling pitches).

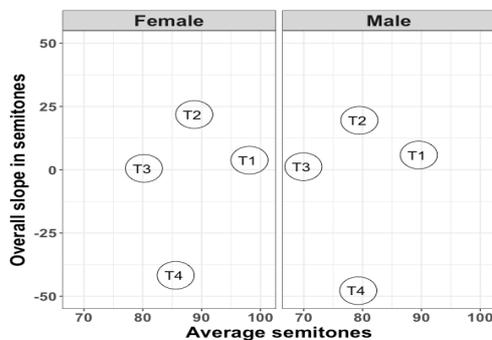


Figure 3: Acoustic space of test stimuli.

Individual Differences Scaling (INDSCAL) was used to examine the perceptual space. This method assumes that the perceptual distance between stimuli can be discerned based on the time taken to discriminate between sounds. For instance, stimuli that are perceptually dissimilar result in shorter reaction times while stimuli that are perceptually similar result in longer reaction times. The output of the INDSCAL multidimensional scaling analysis results in a group stimulus space. This serves as the representation of the four tones in Euclidian space, which can then be interpreted without the need for reconfiguration [5]. Each participant's weighting pattern can further be statistically analyzed. Following the INDSCAL method outlined in [4], a data matrix was created for each participant consisting of his/her distance estimates. These estimates were calculated by taking the normalized inverse of reaction time (1/RT) for each tone-pair comparison (e.g., T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). This resulted in 10 4x4 tone stimulus matrices for native speakers and 30 matrices for the learners (10 participants x 3 testing sessions). INDSCAL analyses of these of dissimilarity matrices were performed at two dimensions following [4,5]. Figure 4 plots native speakers' and learners' configurations at each month of testing.

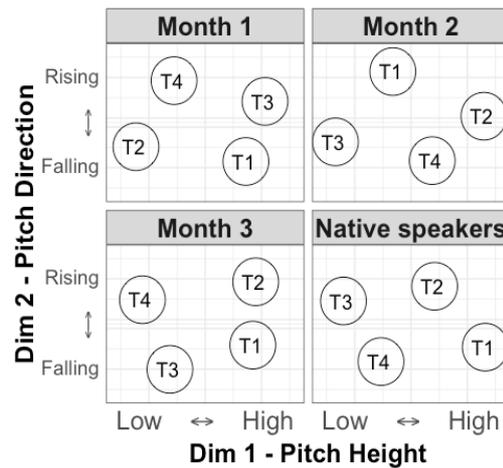


Figure 4: Two-dimensional omnibus INDSCAL analyses

The predicted dimensions of pitch height and pitch direction aligned with the native speakers' configuration in Figure 4. This configuration is similar to that of the acoustic space in Figure 3; tones 2 and 4 are separated largely by pitch direction (dimension 2 along the y-axis) while tones 1 and 3 are separated primarily by pitch height (dimension 1 along the x-axis). The native speakers' MDS output corroborates previous MDS studies e.g., [2,4,6,17]. For the learners, results from months 1, 2, and 3 all show different configurations, none of which fully resemble that of native speakers or that of the acoustic space in Figure 3. In month 1, learners struggled to accurately weight cues in either dimension, resulting in a configuration that neither reflects pitch height nor pitch direction. In month 2, learners' configuration shifted to become slightly more native-like; tones 2 and 3 were separated largely by pitch height while tones 1 and 4 were separated by pitch direction. In month 3, learners' configuration shifted again. This configuration ostensibly reflects learners' attention to pitch end points: tones 1 and 2 end with a high pitch while tones 3 and 4 end with a low pitch. The pitch direction dimension reflects tone 2's rising pitch but not tone 3's rising pitch or tone 4's falling pitch. A one-way ANOVA of the mean participant weights at each month of testing was conducted on each of the two dimensions. For dimension 1 (pitch height), no difference was observed across the three months of testing ($F(2,28) = 1.2$, $p > .05$). For dimension 2 (pitch direction), a marginal effect of month was found ($F(2,28) = 2.7$, $p = .07$). Further analyses of means indicated an increase in weight of the pitch direction dimension as a function of time spent studying.

3. Discussion

The present study examined how Mandarin L2 learners change the weights of dimensions underlying their perceptual space during the first three months of classroom learning. Three analyses were carried out on speeded AX-discrimination data. First, d-prime as a measure of sensitivity was calculated (Figure 1). Learners' d-prime values increased from month 1 to month 2, though this increase was not statistically significant. From month 2 to month 3, learners' sensitivity plateaued. At month 3, learners' mean d-prime was still significantly lower than that of native speakers. Reaction time data followed a similar trend (Figure 2). Learners responded significantly faster in month 2 as compared to

month 1 but showed near identical reaction times in months 2 and 3. Month 3 reaction times were marginally different from those of native speakers, indicating that after three months of classroom learning non-native listeners' correct discrimination speeds were roughly comparable to those of native listeners. INDSICAL analyses of the AX-discrimination reaction time data were carried out for natives and at each month for learners. Two dimensions were identified: pitch height and pitch direction (Figure 4), which corresponded to the acoustic space of the test stimuli (Figure 3). Analyses of the learners' mean weights indicated that the pitch direction dimension changed at a marginal rate as learners gradually attended more to pitch direction dimensions and less to pitch height dimensions. Yet, learners' configurations at months 2 and 3 did not fully resemble that of native speakers. In particular, month 3's configuration suggests that learners attended to pitch change but perseverated on the end points' height rather than the overall pitch direction or contour. While tones 3 and 4 both end with lower pitch heights, these tones' respective pitch contours critically differ such that pitch direction (dimension 2) is more informative for categorization. This implies an intermediate step for learners between weighting pitch height dimensions and weighting pitch direction dimensions: learners in month 3 attended to the pitch height at the end of the contour rather than the overall pitch direction (Figure 4). This continued emphasis on pitch height dimensions seems to be the cause of the learning plateau observed in d-prime sensitivity and reaction time.

Is it the case that all participants in the present study experienced this learning plateau or was it simply a group of what [4] refers to as "poor" learners? Previous work has shown that "good" learners of Mandarin do not necessarily have better pitch perception abilities, but rather these learners place more emphasis on pitch direction [3,4,7,8]. This finding was replicated in the present study: learners shifted their cue-weighting of pitch direction across the three months of learning at a marginal rate. Individual analyses of learners' matrices confirmed that mean dimension 2 (pitch direction) weights increased from month 1 to month 3 in the same direction for all ten participants. The large individual variability seen in [4] was not observed in the present study: all participants completed the AX discrimination task with above 80% accuracy at month 1 and above 85% accuracy at months 2 and 3. Mean reaction times and d-prime values demonstrated consistently small standard errors across participants. Moreover, while results from the present study cannot be directly compared to results from [4] (the two studies used different stimuli), the ten learners tested in the present study had mean reaction times and d-prime results equal to (if not better than) those of the "good" learners in [4]. Thus, even "good" learners who attend to pitch direction and are motivated to learn the language in a classroom setting can still encounter a learning plateau during early L2 acquisition.

Given the present findings, it is possible that the gains observed in terms of improved d-prime sensitivity and faster reaction time from month 1 to month 2 may be the same initial gains observed in typical perceptual learning studies, e.g., [4,18]. Laboratory-based perceptual learning paradigms provide listeners with short, input intensive training sessions that lead to relatively rapid improvements in categorization. Yet, the gains observed in these perceptual learning studies are difficult to reconcile with consistent findings that classroom learners encounter a learning plateau after which they struggle to make further gains [11]. This is especially true in foreign

languages like Chinese, which generally require English L1 speakers longer to master than European languages like Spanish or French [18,11]. The present results suggest that laboratory-based perceptual learning exercises may not be sufficient in moving early learners off the learning plateau. Recent approaches to perceptual learning of non-native sounds argue that incidental learning, in which a learner's attention is directed away from the stimuli to be learned by engaging in a task distinct from the L2 learning domain, can lead to efficient auditory and speech category learning [19,20]. Future work will need to evaluate the success of incidental learning on long term L2 lexical tone acquisition.

In sum, the present study's data are consistent with previous studies on Mandarin tone learning and perception by non-native listeners, e.g., [3,4,7,8,18]. Early classroom learners make initial gains in tone categorization by attending more to pitch direction dimensions and less to pitch height dimensions. Even motivated, "good" learners, however, encounter a learning plateau in their cue-weighting abilities. New perceptual learning and L2 training paradigms are needed to help advance learners off the learning plateau.

4. References

- [1] P. Iverson, P. K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Ketemann, and C. Siebert, "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition*, vol. 87, pp. B47-B57, 2003.
- [2] J. T. Gandour, "Tone perception in Far Eastern languages," *Journal of Phonetics*, vol. 11, pp. 149-175, 1983.
- [3] B. Chandrasekaran, J. T. Gandour, and A. Krishnan, "Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity," *Restorative Neurology and Neuroscience*, vol. 25, pp. 195-210, 2007.
- [4] B. Chandrasekaran, P. D. Sampath, and P. C. M. Wong, "Individual variability in cue-weighting and lexical tone learning," *Journal of Acoustic Society of America*, vol. 128, no. 1, pp. 456-465, 2010.
- [5] J.D. Carroll and P. Arabie, "Multidimensional scaling," *Annual Review of Psychology*, vol. 31, pp. 607-649, 1980.
- [6] A. L. Francis, V. Ciocca, L. Ma, and K. Fenn, "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," *Journal of Phonetics*, vol. 36, pp. 268-294, 2008.
- [7] P. C. M. Wong and T. K. Perrachione, "Learning pitch patterns in lexical identification by native English-speaking adults," *Applied Psycholinguistics*, vol. 28, no. 4, pp. 565-585, 2007.
- [8] T. K. Perrachione, J. Lee, L. Y. Y. Ha, and P. C. M. Wong, "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," *The Journal of The Acoustical Society of America*, vol. 130, pp. 461-472, 2011.
- [9] K. Csizér and Z. Dörnyei, Z., "The internal structure of language learning motivation and its relationship with language choice and learning effort," *The Modern Language Journal*, vol. 89, pp. 19-36, 2005.
- [10] R. C. Gardner, *Social psychology and second language learning: The role of attitudes and motivation*, Arnold Publishing, 1985
- [11] L. Colantoni, J. Steele, and P. Escudero, P., *Second Language Speech*, Cambridge University Press, 2015.
- [12] G. R. Slavoff, and J. S. Johnson. "The effects of age on the rate of learning a second language." *Studies in Second Language Acquisition*, vol. 17, pp. 1-16, 1995.
- [13] Y. Xu, Y., "Contextual tonal variations in Mandarin," *Journal of Phonetics*, vol. 25, pp. 61-83, 1997.
- [14] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 5.3.62. <http://www.praat.org/>, 2015.

- [15] D. Bates, M. Maechler, and B. Bolker, *lme4: Linear mixed-effect models using eigen and classes*, <http://cran.r-project.org/package=lme4>
- [16] U. Halekoh, and S. Højsgaard, “A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest,” *Journal of Statistical Software*, vol. 59, pp. 1-30, 2014.
- [17] J. T. Gandour, “Perceived dimensions of 13 tones: A multidimensional scaling investigation,” *Phonetica*, vol. 35, pp. 169-179, 1978.
- [18] Y. Wang, M. Spence, A. Jongman and J. Sereno, “Training American listeners to perceive Mandarin tones,” *Journal of Acoustic Society of America*, vol. 106, no. 6, pp. 3649-3658, 1999.
- [19] T. Wade, and L. L. Holt, “Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task,” *The Journal of the Acoustical Society of America*, vol. 118, pp. 2618-2633, 2005.
- [20] R. Liu, and L. L. Holt, “Neural changes associated with non-speech auditory category learning parallel those of speech category acquisition,” *Journal of Cognitive Neuroscience*, vol. 23, pp. 683–698, 2011.