Incidental learning of non-speech auditory analogs scaffolds second language learners' perception and production of Mandarin lexical tones

Seth Wiener, Timothy K. Murphy, Atul Goel, Michael G. Christel, Lori L. Holt

Carnegie Mellon University
sethw1@cmu.edu, tkmurphy@andrew.cmu.edu, atulg@andrew.cmu.edu, christel@cmu.edu, loriholt@cmu.edu

ABSTRACT

Mandarin Chinese lexical tones are notoriously difficult for second language (L2) learners to accurately perceive and produce. In Experiment 1, we demonstrate how adult classroom L2 Mandarin learners typically plateau in their tone learning abilities. In Experiment 2, we train a new group of classroom L2 learners on non-speech auditory analogs of Mandarin tone categories. We make use of incidental learning in which learners' attention is directed away from the tone categories by an engaging videogame. Results indicate that non-speech incidental learners improved their categorization of Mandarin tone relative to classroom learners who were explicitly trained on Mandarin speech. Moreover, incidental learning transferred to affect learners' reading aloud of tones, resulting in more native-like tonal contours. Non-speech 'perceptual building block' categories that share critical perceptual features with L2 speech thus appear to support classroom learning of difficult-to-acquire L2 speech sounds.

Keywords: Incidental learning, Mandarin Chinese, lexical tone, second language acquisition

1. INTRODUCTION

Mandarin Chinese (hereafter 'Mandarin') lexical tones are often cited as a paradigmatic example of difficult-to-acquire non-native speech sounds. For native speakers of a non-tonal language such as English, learning to accurately perceive and produce the four pitch patterns is notoriously difficult [5-7]. For many adult second language (L2) Mandarin learners, perceptual and articulatory abilities plateau during classroom learning; even after multiple years of classroom learning, L2 learners fail to approach native-like abilities [5, 18]. One possible explanation for these challenges is that tone learning may be constrained by explicit instruction of the Mandarin tones' acoustic-phonetic characteristics (e.g., F0 contours, duration, amplitude [4, 16]). Whereas classroom tone learning can lead to modest L2 gains, explicit learning may interact unproductively with first language (L1) prosodic categories, which have some overlapping characteristics with Mandarin tone [14]. Recent research has found that, quite counter-intuitively, 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 [9-10]. To date, however, L2 incidental learning studies have primarily taken a lab-based approach in line with foundational perceptual learning studies on segmental contrasts, e.g., [2]. It is possible that the observed gains from short-term incidental learning tasks may be comparable to (or interfere with) those observed within a structured L2 classroom. We therefore extend incidental tone learning research to test adults engaged in classroom L2 Mandarin learning. In doing so, we make use of a particular type of incidental learning, which utilizes non-speech auditory analogs of Mandarin tone categories. Previous research has found that building a foundation of nonlinguistic 'perceptual building block' categories results in generalization of learning to the L2 speech categories that the nonlinguistic sounds model [9, 15]. Importantly, these building blocks share critical perceptual dimensions with tone categories, but are not perceived as speech and may therefore provide a 'back door' through which to influence adult L2 learners' speech acquisition.

We present preliminary results from two experiments. In Experiment 1, we replicate previous L2 tone learning results, e.g., [5, 6, 15], that highlight the acquisition challenges facing typical classroom L2 Mandarin learners. In Experiment 2, we train a new set of classroom learners using non-speech auditory analogs of Mandarin tone embedded in a novel incidental learning videogame. These results are compared to control learners who train explicitly through Mandarin speech learning tasks. We demonstrate how learners incidentally trained on non-speech perceptual building blocks of tone improve in their categorization and production of tone to a greater degree than those trained explicitly on Mandarin speech. Taken together, these results suggest that incidental videogame training on non-speech perceptual categories that align with perceptual dimensions of L2 categories can scaffold wider L2 success in both speech perception and speech production.

2. EXPERIMENT 1

2.1. Methods

Eight adult native-English university students participated in Experiment 1 as the L2 Mandarin group. All participants had normal hearing, spoke only English, and were enrolled in a first-semester Mandarin language course, which met each week for 3.5 hours of classroom instruction. Participants' musical background and pitch perception were controlled. Eight native-Mandarin university students served as the L1 group. All participants self-reported only speaking Mandarin, i.e., no other tonal dialect, and spoke English as an L2. All L1 English participants reported in this paper were paid or given class credit for their time. All L1 Mandarin participants volunteered their time.

2.1.1. Stimulus creation

Stimuli creation followed the methods in [3, 20]: the Mandarin vowel /y/ with all four citation-form F0 contours was recorded by two native speakers (one female, one male) and sampled at 44.1 kHz. Praat's pitch synchronous overlap and add (PSOLA) method [1] was used to superimpose each F0 contour on the vowel. Stimuli had a normalized duration of 400 ms (female) and 450 ms (male).

2.1.2. Procedure

Participants performed two perception tasks: a pitch contour discrimination sensitivity task (AX discrimination) and a four-alternative forced-choice (4AFC) tone categorization task. In the AX task, participants were told to identify whether they heard the "same" or "different" speech sounds as quickly and accurately as possible. Participants were given a short practice session, followed by 196 trials across two blocks. In each trial, the two sounds were always spoken by the same speaker with a 500 ms ISI and a 1 second timeout period. Trials were counterbalanced for an equal number of same/different and male/female trials. Sensitivity index (d') was calculated for each participant.

In the 4AFC task, participants were shown the four Mandarin tone contours on screen and told to identify the pitch pattern of each sound as quickly and accurately as possible. Participants were given a short practice session, followed by 196 trials across two blocks with a 1 second timeout period. Trials were counterbalanced for an equal number of each tone type and male/female utterances.

Participants also performed a tone production reading task. Participants were asked to read aloud a list of 20 familiar words written in Pinyin romanization (e.g., *hǎo*; 5 words per tone type). Recordings were presented to 5 L1 Mandarin raters who were asked to identify the perceived tone category of each utterance. Tone accuracy therefore reflects whether the native speaker perceived the utterance as the intended tone category.

The L2 group was tested three times after roughly 1, 2, and 3 months of classroom learning. During the first two testing sessions, participants only performed the two perception tasks (counterbalanced in presentation order). At the third testing session, participants performed the reading task in addition to the two perception tasks. The L1 group was only tested once and performed all three tasks. All testing occurred in a quiet lab over headphones and lasted approximately 20 minutes.

2.2. Results

Ninety-five percent confidence intervals (CIs) for mean d', 4AFC accuracy, and reading accuracy were calculated for each group and reported in Table 1.

Table 1: Experiment 1 95% confidence intervals for mean d', 4AFC accuracy, and tone reading.

	d'	4AFC	Reading
L2 month 1	[1.9, 2.8]	[.77, .90]	
L2 month 2	[3.1, 3.4]	[.91, .94]	
L2 month 3	[3.2, 3.6]	[.91, .94]	[.34, .58]
L1 Mandarin	[5.1, 5.8]	[.99, 1]	[.99, 1]

Analyses of d' revealed a main effect of group, $F(3, 28) = 101.7$, $p < .001$, $\eta^2 = .916$. Post-hoc comparisons revealed that the L1 group had a higher mean d' than the L2 group at all three tests (ps < .05). The L2 group had a higher mean d' at months 2 and 3 compared to month 1 (ps < .05); no difference was found between months 2 and 3 ($p > .05$).

Analyses of 4AFC accuracy revealed a main effect of group, $F(3, 28)=24.81$, $p < .001$, $\eta^2 = .726$. Post-hoc comparisons revealed that the L1 group was more accurate than the L2 group at all three tests (ps < .05). The L2 group was more accurate at months 2 and 3 compared to month 1 (ps < .05); no accuracy difference was found between months 2 and 3 ($p > .05$).

The L1 group was more accurate than the L2 group at reading aloud the Pinyin tones, $t(7) = -24.41$, $p < .001$, $d = 12.2$.

The results from Experiment 1 corroborate previous findings on L2 tone learning, e.g., [5]. Learners' sensitivity and categorization of tone improved during the first two months of structured L2 classroom learning. Yet, these initial gains were followed by a learning plateau in which fewer gains were observed. In Experiment 2 we make use of

incidental learning with non-speech auditory analogs as a potential means to advance learners beyond this persistent learning plateau.

3. EXPERIMENT 2

3.1. Methods

Eight new L2 Mandarin learners participated in Experiment 2 as the L2 group. All participants met the same recruitment criteria used in Experiment 1 and were enrolled in a different section of the same first-semester Mandarin language course.

3.1.1. Procedure

The tasks and materials followed those of Experiment 1 with the expectation that participants were only tested at months 1 and 3. Between months 1 and 3, participants took part in either incidental non-speech training (N = 4) or explicit speech training (N = 4) for 6-weeks (approximately 30/min a week).

Explicit speech training involved web-based Mandarin tone discrimination, categorization, and labelling tasks from the participants' L2 Mandarin textbook. For example, participants heard a recorded utterance of *ma* and had to indicate whether the speaker said *má* or *mǎ.* These speech-learning exercises followed an audio-lingual approach in line with current L2 Mandarin pedagogy, e.g., [21]. Participants performed the exercises in a quiet space using headphones and a computer.

Incidental non-speech training involved a space themed videogame in which alien spaceships appeared on screen while a non-speech hum sound was played [8]. These nonlinguistic sounds were impossible vocalizations that lacked information about phonetic segments and voice but which mimicked the pitch contours typical of Mandarin tone categories and were derived such that they perfectly modeled the perceptual space characterizing four different Mandarin talkers' lexical tone contours (see [15] for additional details regarding auditory analogs). Due to the nature of the task, participants were neither explicitly instructed to form audio-visual or audio-motor associations, nor were they told the significance of the sounds. However, a few game mechanics strongly promoted auditory category learning. Firstly, each spaceship was associated with a particular sound category. Each time a ship appeared, a randomly-selected, acoustically variable sound exemplar drawn from an associated sound category was presented until the participant aimed the laser and executed an action. Secondly, each ship originated from a consistent quadrant of visual space (with some jitter). Players had a limited shooting range at any given point that roughly spans the potential range of approach for any one of the spaceships. This made it possible to "set" the laser's range even before a ship appears, but only if participants identified the upcoming ship's quadrant using the sound category that predicts the ship. Thirdly, as the game progressed, the speed and difficulty of the game increased so that quick identification of approaching ships by their sound category (while never required or explicitly encouraged) became of gradually increasing benefit. Participants played the videogame in a quiet space using headphones and a computer.

3.2. Results

Ninety-five percent CIs for mean d', 4AFC accuracy, and reading accuracy were calculated for each group and reported in Table 2.

Table 2: Experiment 2 95% confidence intervals for mean d', 4AFC accuracy, and tone reading.

	d'	4AFC	Reading
Explicit			
+speech			
Month 1	[1.5, 3.0]	[.70, .93]	
Month 3	[2.7, 4.2]	[.92, .95]	[.29, .50]
Incidental			
+non-speech			
Month 1	[1.6, 2.9]	[.73, .92]	
Month 3	[3.2, 4.6]	[.95, .98]	[.44, .57]

Analyses of the two L2 groups' d' results revealed no difference at month 1, $t(6) = -0.05$, $p = .96$, and resembled those of the L2 group at month 1 in Experiment 1, $F(2, 13) = 0.08$, $p = .91$. The two L2 groups' mean d' at month 3 did not differ, $t(6) = -1.52$, $p = .17$. Both groups showed a significant d' improvement between month 1 and 3 (ps $< .05$); both groups resembled the L2 group tested in Experiment 1 at month 3, $F(2, 13) = 3.20$, $p = .07$.

Analyses of the two groups' 4AFC accuracy at month 1 revealed no difference, $t(6) = -0.10$, $p = .91$, and resembled that of the L2 group tested at month 1 in Experiment 1, $F(2,13) = 0.06$, $p = .94$. At month 3, however, the incidental non-speech group was more accurate than the explicit speech group, $t(6) = -3.73$, $p < .01$, $d = 2.64$, though both groups showed a significant improvement between month 1 and 3 (ps $< .05$). Post-hoc comparisons revealed the explicit speech group performed similarly to the participants tested in Experiment 1 ($p > .05$) whereas the incidental non-speech group outperformed Experiment 1 participants ($p < .05$).

A reading accuracy difference was found, as the incidental non-speech group read aloud tones more accurately than the explicit speech group, $t(6)=-2.97$, $p < .05$, $d = 2.10$. The explicit speech group and participants in Experiment 1 did not differ in mean tone reading accuracy ($p > .05$).

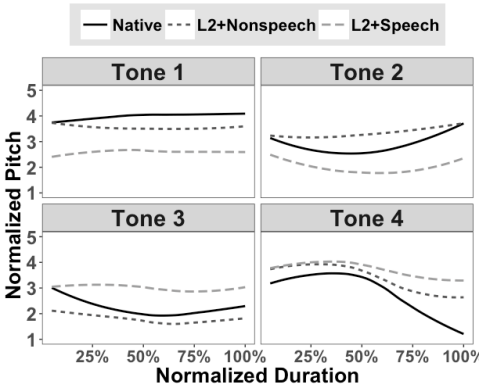
4. DISCUSSION

Acquiring Mandarin lexical tones as an L2 learner can be incredibly frustrating. As a result, many L2 learners abandon classroom learning before reaching a high level of Mandarin proficiency [7]. In Experiment 1 we demonstrated that while classroom L2 learners make initial gains in tone discrimination, categorization, and tone reading, learners' typically plateau after several months of explicit classroom instruction.

In Experiment 2 we explored whether supplementary explicit speech training or incidental non-speech training could support wider L2 Mandarin acquisition. We found that 6 weeks (roughly 30 min/week) of explicit speech training did not improve learners' perception or production of L2 tones beyond the levels attained without the intervention. In contrast, we found that 6 weeks of supplementary incidental training using non-speech auditory analogs of Mandarin tone resulted in improved categorization and more native-like Mandarin tone productions.

Taken together, the present study's preliminary results suggest that nonlinguistic 'perceptual building block' categories result in generalization of learning to the L2 speech categories that the nonlinguistic sounds model, e.g., [9, 10]. Because these building blocks share critical perceptual dimensions with tone categories, we argue that they provide a 'back door' through which to influence adult L2 learners' speech acquisition without the interference of explicit knowledge of L1 categories. Novel to the present study, we extended previous lab-based research to adults engaged in structured L2 classroom learning. To our knowledge, this serves as initial evidence of incidental learning supporting wider natural adult L2 acquisition. Even more exciting for future research and theoretical models of L2 speech learning, we observed that incidental non-speech training on auditory analogs of tone transferred to affect L2 learners' articulation of acoustic-phonetic cues involved in accurate tone productions. This transfer occurred despite learners receiving no additional production practice during the incidental training. Figure 1 plots the normalized pitch contours of the L2 learners in Experiment 2 and the native speakers in Experiment 1 (following the normalization methods outlined in [17]). This figure shows that learners trained on non-speech building blocks produced tone contours that approximated those of native Mandarin speakers whereas those trained explicitly on Mandarin speech produced less native-like contours.

Figure 1: Normalized pitch contours for L2 learners' productions in Experiment 2 and L1 Mandarin speakers' productions in Experiment 1.



We acknowledge that the present study's preliminary results are based on a small sample size and restricted to classroom L2 learners who may not reflect the typical adult L2 learner. Moreover, although our incidental training program improved classroom learners' abilities, we note that participants remained considerably less accurate than native speakers, as well as more proficient L2 learners tested in prior research, e.g., [13,16]. Future research is needed to strengthen our claims and clarify whether extended incidental training may lead to even greater L2 improvement. Additionally, because non-lexical, acoustic-phonetic tone processing (as emphasized in our AX and 4AFC tasks) differs in important ways from the *lexical* tone processing necessary for Mandarin language acquisition [11], we recognize that our results may reflect task-specific perceptual strategies rather than the processes and mechanisms supporting Mandarin tonal word learning [12, 19]. For these reasons, our future work will expand outcome metrics beyond perceptual categorization and generalization.

5. ACKNOWLEDGMENTS

This work was supported by an A. W. Mellon ProSeed Grant through Carnegie Mellon University.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 5.3.62. http://www.praat.org/

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

[3] Chandrasekaran, B., Sampath, P. D., Wong, P. C. M. 2010. Individual variability in cue-weighting and lexical tone learning. *J. Acoust. Soc. Am. 128*, 456-465.

[4] Gandour, J. T. 1983. Tone perception in Far Eastern Languages. *J. Phonetics, 11*, 149-175.

[5] Hao, Y. C. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *J. Phonetics*, *40*(2), 269-279.

[6] Hao, Y. C. 2018. Second language perception of Mandarin vowels and tones. *Lang. Speech*, *61*(1), 135-152.

[7] Ke, C., Reed, D. J. 1995. An analysis of results from the ACTFL Oral Proficiency Interview and the Chinese Proficiency Test before and after intensive instruction in Chinese as a foreign language. *Foreign Lang. Annals 28*(2), 208-222.

[8] Kimball, G., Cano, R., Feng, J., Feng, L., Hampson, E., Li, E., Christel, M. G., Holt, L. L., Lim, S.-J., Liu, R., Lehet, M. 2013. Supporting research into sound and speech learning through a configurable computer game *Proc. IEEE Games Innovation Conf.*, 110-113.

[9] Lim, S. J., Holt, L. L. 2011. Learning foreign sounds in an alien world: videogame training improves non-native speech categorization. *Cog. Sci. 35*(7), 1390–1405.

[10] Liu, R., Holt, L. L. 2011. Neural changes associated with non-speech auditory category learning parallel those of speech category acquisition. *J. Cog. Neuroscience, 23*, 683–698.

[11] Malins, J. G., Joanisse, M. F. 2012. Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*, 2032-2043.

[12] 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. *J. Acoust. Soc. Am. 130*, 461-472.

[13] Shih, C., Liu, H. Y. D. 2015. Effects of talker-to-listener distance on tone. *J. Phonetics, 51*, 6-35.

[14] So, C. K., Best, C. T. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Lang. Speech, 53*(2), 273-293.

[15] Wade, T., Holt, L. L. 2005. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J. Acoust. Soc. Am. 118*, 2618-2633.

[16] Wang, Y., Spence, M., Jongman A., Sereno, J. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am. 106*, 3649-3658.

[17] Wang, Y., Jongman, A., Sereno, J. A. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc Am., 113*, 1033-1043.

[18] Wiener, S. 2017. Changes in Early L2 Cue-Weighting of Non-Native Speech: Evidence from Learners of Mandarin Chinese. *Proc. Interspeech* Stockholm, 1765-1769.

[19] Wiener, S., Ito, K., Speer, S. R. 2018. Early L2 Spoken Word Recognition Combines Input-Based and Knowledge-Based Processing. *Lang. Speech*, *61*(4), 632-656.

[20] Xu, Y. 1997. Contextual tonal variations in Mandarin. *J. Phonetics, 25*, 61-83.

[21] Zhang, H. 2018. *Second Language Acquisition of Chinese Tones—Beyond First-Language Transfer,* Brill.