The Thesis Committee for Rachel Sheridan Tessmer
Certifies that this is the approved version of the following thesis:

Optimizing Lexical Learning by Manipulating Phonological Training

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Bharath Chandrasekaran

Madhu Sundarrajan

Optimizing Lexical Learning by Manipulating Phonological Training

by

Rachel Sheridan Tessmer, B.S.Psy

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2017

Acknowledgements

This experiment was conducted in conjunction with Han-Gyol Yi and Dr. Bharath Chandrasekaran. This research was funded by NIH-NIDCD grant R01DC013315 awarded to BC and an Undergraduate Research Fellowship awarded to RT. Han-Gyol Yi provided invaluable assistance in every step of the process. Without his insight, this would not have been possible. I want to especially thank Dr. Bharath Chandrasekaran for his mentorship, guidance, and support throughout the duration of this project. Thank you to the research assistants in the SoundBrain Laboratory for assistance with participant recruitment and data collection. Thanks to Zilong Xie and Bethany Hamilton for their input regarding data visualization. I would also like to thank Dr. Bharath Chandrasekaran and Dr. Madhu Sundarrajan for their helpful feedback.

Abstract

Optimizing Lexical Learning by Manipulating Phonological Training

Rachel Sheridan Tessmer, M.A.

The University of Texas at Austin, 2017

Supervisor: Bharath Chandrasekaran

In this thesis, I examine the impact of manipulating phonological training environments on learning novel words in adulthood. Adults who learn non-native phonological categories are thought to utilize two main types of strategies. A *reflective* strategy involves developing explicit rules for categorizing sounds, whereas a *reflexive* strategy entails implicitly mapping sounds onto representations. Successful learning has been associated with a transition from a reflective to a reflexive strategy. This study examined the extent to which successful phonological category learning can lead to the enhanced acquisition of novel lexical items based on the same phonological category structure. Monolingual English speakers (N=40) learned to categorize Mandarin tones in one of four training regimens. Participants in a Reflect→Reflex condition were initially presented with a training environment designed to enhance reflective learning and then with an environment designed to enhance reflexive learning. Participants in a Reflex→Reflect condition were presented with the two environments in a reversed order. Training environments in two control conditions exclusively targeted either reflective or reflexive learning. Following phonological training, participants were trained to identify novel pseudo-Mandarin lexical items using a sound-to-meaning training paradigm across three days. Participants in the Reflect→Reflex condition outperformed all conditions on the lexical task, despite equivalent performance in the earlier phonological categorization task. Results support initially targeting reflective and later targeting reflexive systems during speech category learning. The findings suggest that reducing challenges in phonological categorization via optimized training approaches may help bootstrap novel word learning.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Learning a second language during adulthood is often a challenging experience. This challenge has frequently been attributed to the 'critical period' (Penfield & Roberts, 1959; Lenneberg, 1967; Werker & Hensch, 2015). The critical period hypothesis posits that the capability to learn language reaches a peak earlier in life followed by a decline with increasing age. Adults attempting to learn a second language may find greater difficulty reaching native-like levels of proficiency and fluency than children (Johnson & Newport, 1991). This difficulty is in part due to challenges in perceiving non-native speech sounds (Wang, Spence, Jongman, & Sereno, 1999). Adult language learners are less sensitive to important acoustic dimensions that differentiate non-native speech sounds, relative to infant language learners (Kuhl, 1983; Iverson et al., 2003). It may also be the case that the phonetic system of adult language learners' native language influences their ability to create separate categories for more similar non-native speech sounds (Flege, 1995; Best, 1995; Wang et al., 1999). This account suggests that adult language learners perceive non-native sounds with reference to categories of their existing language system, and that perceived phonetic similarity may impact how well non-native sounds are acquired.

In this thesis, I will focus on native English adult learners acquiring phonological categories in Mandarin Chinese. Mandarin is a tonal language in which changes in pitch categories at the syllable level impact the meaning of words (Song, Skoe, Wong, & Kraus, 2008). The four lexical pitch contrasts are described by their pitch height and direction patterns as tone 1 (T1: high-level), tone 2 (T2: low-rising), tone 3 (T3: low-dipping), and tone 4 (T4: high-falling) (Wong & Perrachione, 2007; Chandrasekaran, Sampath, & Wong, 2010; Chandrasekaran, Koslov, & Maddox, 2014). An example of this is that the syllable /ma/ produced with T1 means 'mother', while /ma/ produced with T2 means 'hemp', /ma/ produced with T3 means 'horse', and /ma/ produced with T4 means 'scold' (Chao, 1968). This particular phonemic contrast does not occur in English. The pitch differences of these tones present a particular obstacle for native English speakers trying to learn Mandarin as adults (Wang, Jongman, & Sereno, 2003; Wang et al., 1999; Kiriloff, 1969). Learners must map these multidimensional signals into distinct and meaningful categories (Lister, 1986; Hillenbrand, Getty, Clark, & Wheeler, 1995; Jongman & Moore, 2000; Vallabha, McClelland, Pons, Werker, & Amano, 2007; Holt & Lotto, 2008; Holt & Lotto, 2010).

NON-NATIVE SPEECH CATEGORY TRAINING

With sufficient experience and exposure, adult language learners can learn to distinguish non-native speech sounds that do not have phonemic counterparts in their native language; however, it remains difficult to form separate categories for more similar sounds (Flege, 1987; Best & Jones, 1988; Wang et al., 1999). Non-native speech sound categories can be learned with training. Holt and Lotto (2008; 2010) have likened the task of mapping variable speech signals from multiple talkers to specific categories to a categorization problem. Evidence exists for language-specific speech categories (Näätänen et al., 1997). Possible influences from a learner's native speech categories or perceptual warping from prior language experience may make it more challenging to learn non-native speech categories (Best, 1993; Best, Morrongiello, & Robson, 1981; Best & Tyler, 2007; Flege, Yeni-Komshian, & Liu, 1999; Francis, Ciocca, Ma, & Fenn, 2008; Francis & Nusbaum, 2002).

Even with potential influences from native speech categories and prior experience, research has shown that English speakers who have no prior exposure to tonal languages can be trained to accurately categorize Mandarin tones after several training sessions (Chandrasekaran, Koslov, & Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Chandrasekaran, Yi, Blanco, McGeary, & Maddox, 2015; Wang et al., 1999; Maddox, Chandrasekaran, Smayda, & Yi, 2013; Yi, Maddox, Mumford, & Chandrasekaran, 2016). Two dimensions that are important for distinguishing Mandarin tones are pitch direction and height. The perceptual saliency of these dimensions can be influenced by the tonal rules and patterns in a learner's native phonological system (Gandour, 1978; Hume & Johnson, 2001; Maddox & Chandrasekaran, 2014). Krishnan and colleagues (2005) found that native speakers of Mandarin Chinese showed stronger pitch representation, stronger second harmonic representation, and better pitch tracking in their frequency-following response recordings relative to English speakers. Other studies have found that pitch direction is weighted more by native speakers of tonal languages (Francis et al., 2008; Chandrasekaran, Gandour, & Krishnan, 2007; Massaro, Cohen, & Tseng, 1985; Gandour & Harshman, 1978). However, some of these differences can be reduced following training. For instance, Song and colleagues (2008) examined the frequency following response of native English speakers prior to and following sound-to-meaning training on lexical pitch patterns in a word identification task. After training, native English speakers had more accurate pitch tracking with fewer pitch-tracking errors. Chandrasekaran and colleagues (2010) found that attending to pitch direction led to increased accuracy on sound-to-meaning auditory tasks for English speakers. Another study found that explicitly instructing native English speakers to attend to the pitch direction dimension led to greater tone categorization accuracy (Chandrasekaran et al., 2016).

Feedback-based categorization tasks are designed to maximize generalization of learning. In laboratory-based training studies, training environments that utilize natural stimuli produced by multiple talkers and provide trial-by-trial feedback have been found to be useful (Lively et al., 1994; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006; Zhang et al., 2009; Maddox & Chandrasekaran, 2014; Lim & Holt, 2011; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Yi et al., 2016; Chandrasekaran, Yi, & Maddox, 2014; Chandrasekaran et al., 2015). The use of multiple talkers during training helps learners adjust their categories by incorporating various productions and helps learners attend to talker-invariant acoustic cues that are important for distinguishing between categories (Bradlow, 2008). Multi-talker training has been shown to lead to better generalization, suggesting that variability during phonological training is beneficial for some learners (Lively et al., 1994; Bradlow, 2008; Lively, Logan, & Pisoni, 1993). In addition, providing native English speakers with trial-by-trial feedback during speech categorization training has been shown to lead to greater learning gains than unsupervised learning paradigms (McClelland, Fiez, & McCandliss, 2002; Vallabha & McClelland, 2007; Yi et al., 2014). Feedback allows for continuous error monitoring.

INDIVIDUAL DIFFERENCES IN NON-NATIVE SPEECH CATEGORY LEARNING

There is a great deal of variability in learning success across non-native speech categorization training studies. While some learners learn categories quickly, others take longer to reach high accuracy levels, and others still have difficulty reaching accuracy beyond chance performance (Chandrasekaran et al., 2015). Perceptual differences may be one source of individual variability in speech learning success (Perrachione, Lee, Ha, & Wong, 2011). Various studies have found that individual differences in auditory regions prior to training and differences in functional connectivity in frontal and parietal brain regions contribute to individual variability in non-native speech learning success (Maddox & Chandrasekaran, 2014; Wong, Perrachione & Parrish, 2007; Golestani, Molko, Dehaene, LeBihan & Pallier, 2007; Wong, Chandrasekaran, Garibaldi & Wong, 2011; Ventura-Campos et al., 2013). Another proposed source of this variability is that individuals use different learning strategies. Behavioral studies, neuroscience studies, and computational modeling all point towards the existence of at least two, competitive learning systems (dual-learning systems) involved in category learning (Nomura & Reber, 2008; Knowlton, 1999; Waldron & Ashby, 2001; Chandrasekaran, Koslov, & Maddox, 2014).

The dual-learning systems model posits that there are two competing systems where learning occurs: a *reflective* system and a *reflexive* system (Chandrasekaran, Koslov, & Maddox, 2014). The reflective system is under conscious awareness, involving both executive attention and working memory (Maddox et al., 2013). This learning system is involved in explicit, rule-based learning. Learners test and adjust category rules to determine decision bounds (Chandrasekaran, Koslov, & Maddox, 2014). Learners presumably can test these rules by processing corrective feedback after each trial. The reflective system is mediated by the prefrontal cortex, including the anterior cingulate, anterior caudate nucleus, and dorsolateral prefrontal cortex, as well as by the hippocampus (Chandrasekaran, Yi, & Maddox, 2014; Seger & Miller, 2010; & Ell, 2001). The reflexive system, on the other hand, does not involve working memory. Rather, it involves implicit, procedural learning that associates perceptions with rewarded actions, such as feedback that the learner's category response was correct (Chandrasekaran, Koslov, & Maddox, 2014). This system is mediated by dopamine modulation in the striatum, where cortical-motor responses are associated with cells in sensory association cortex (Chandrasekaran, Koslov, & Maddox, 2014; Ashby & Maddox, 2011. The reflexive system involves the premotor cortex, putamen, and body and tail of the caudate nucleus. Each learning system is thought to be optimal for learning a different type of category structure. The reflective system is thought to be better suited for learning category structures where learners can describe their decisions using clear rules; the reflexive system is thought to be better suited for learning category structureswhere learners may have difficulty verbally describing their process

(Chandrasekaran, Koslov, & Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Maddox & Chandrasekaran, 2014).

The dual-learning systems model predicts that the reflective and reflexive learning systems complement one another in learning different types of category structures, meaning that one system may be more suited for learning one type of category structure over the other (Chandrasekaran, Yi, & Maddox, 2014). In influential dual-learning system models, it is posited that learners will continue to use one system until the other system becomes more accurate (Ashby & Maddox, 2011). Different training paradigms, such as how much information is given when feedback is presented and how talkers are presented, support different learning systems. The dual-learning systems model proposes that both the reflective and reflexive learning systems are feedback-dependent (Chandrasekaran, Yi, & Maddox, 2014). However, these learning systems are dissociable from one another in the dynamics of how feedback and variability in training characteristics impact learning due to the underlying neurobiological differences between the systems. Prior work shows that different types of feedback can enhance speech category learning (McClelland et al., 2002; Chandrasekaran, Yi, & Maddox, 2014). In the real world, learners receive a variety of instructional feedback during auditory learning. Feedback can be rich with information or can provide only minimal information to let the learner know whether they are right or wrong. An example of informationally rich feedback could be a speech-language pathologist implementing explicit rules to guide speech sound articulation during therapy. An example of minimally informative feedback could be a computer beep that indicates that a learner made an incorrect response. On the one hand, informationally rich feedback enhances reflective learning by providing more information that allows for more opportunities to test and reject incorrect hypotheses (Maddox & Chandrasekaran, 2014). On the other hand, minimal feedback enhances reflexive learning by reducing resources available for reflective rumination.

Another area of variability in the training environment is the presence of talker variability. Learners may be exposed to multiple talkers intermixed with each other during training. They may also be exposed to talkers one at a time. Reducing talker variability can promote faster hypothesis testing, which is thought to be less taxing on working memory, and thus supports reflective learning (Chandrasekaran, Yi, & Maddox, 2014). In contrast, randomizing/interleaving talkers in the training environment does not allow learners to predict the next talker, which can disrupt testing rules. Learners are then able to associate talker-invariant acoustic cues with implicit rewards from feedback, which is thought to support reflexive learning.

Based on prior work, specific predictions about whether certain aspects of training environments, such as feedback type and talker presentation, support one learning system over another can be made. In regards to trial-by-trial feedback, informationally rich feedback (also known as full feedback) is thought to enhance reflective learning while minimally informative feedback (also known as minimal feedback) is thought to enhance reflexive learning (Maddox, Love, Glass, & Filoteo, 2008; Chandrasekaran, Yi, & Maddox, 2014). In regards to talker presentation, presenting talkers one at a time (blocked talkers) is thought to enhance reflective learning while randomized (mixed) talkers is thought to enhance reflexive learning. Chandrasekaran and colleagues (2014) found better learning when participants received full feedback compared with minimal feedback. They also found better learning when speaker presentation was mixed compared with blocked. Thus, non-native speech categorization is though to be reflexive-optimal, meaning that it is best learned when learners are exposed to environments that support reflexive learning.

An emerging viewpoint is that the two learning systems are not in fact in competition, but are interactive early in training (Paul & Ashby, 2013). The bootstrap interaction theory posits that successful, early reflective learning can bootstrap and enhance reflexive-optimal task performance. Support exists for a one-way interaction between the systems, wherein the reflective system bootstraps early learning by initially learning the suboptimal reflective strategies and later using the reflexive system to refine learning once it becomes rewarding (Chandrasekaran, Koslov, & Maddox, 2014). Based on the bootstrap interaction theory and the dual-learning systems model, it may be the case that initial reflective strategies and transitioning to reflexive strategies later in training may leverage interactions between the systems for better learning.

RELATIONSHIP BETWEEN NON-NATIVE SPEECH CATEGORY LEARNING AND LEXICAL LEARNING

While the ability to recognize and differentiate non-native speech sounds is critical for learning a second language, word learning is essential for meaningful communication. In addition to other cognitive areas, working memory is critical for language learning, and working memory has been found to predict both native and foreign vocabulary acquisition (Ellis, 1996). Short-term phonological storage plays an important role in long-term vocabulary learning (Papagno, Valentine, & Baddeley, 1991). Vocabulary learning requires learners to hear a sound, hold onto this sound in working memory, and map meaning onto it.

Prior work has been conducted on training native English speakers with no prior tonal language experience to acquire pitch categories (sound-to-category mapping) and then to train them to use these categories to distinguish words (sound-to-word mapping) (see Wong & Perrachione, 2007 and Cooper & Wang, 2013). Contextualized within a phonemic-phonological-lexical continuum, at least two studies have found that engaging in sound-to-category training directly relates to better performance in sound-to-meaning training (Wong & Perrachione, 2007; Cooper & Wang, 2013). Cooper and Wang (2013) found that individuals who undertook tone identification training were better able to identify Cantonese words. This lends support that phonological training can enhance listener perception, which may enhance learning words that are minimally contrasted by tones. In a study using Mandarin tones, Wong and Perrachione (2007) found that listeners who were proficient in identifying tonal pitch patterns were more successful at learning vocabulary items that differed based on lexical tones. This work, in conjunction with prior studies on individual differences, provided the motivation for the present study.

PRESENT STUDY

The current study examines the extent to which optimized non-native speech category training can lead to an enhanced performance in *novel word learning*. To investigate this, four different sound-to-category training paradigms were designed based on the dual-learning systems model. These training paradigms were designed to differentially impact the relative dominance of the reflective and reflexive learning systems. Paradigm manipulations involved the amount of feedback information (full feedback v. minimal feedback) and talker presentation (blocked v. randomized). To date, no study has simultaneously manipulated feedback type and talker presentation during Mandarin tone categorization training. Following phonological categorization training and a categorization generalization task, the impact of sound-to-category training paradigms on sound-to-meaning training for novel pseudowords was assessed by having participants complete five sessions of a lexical training task.

Forty young adult native speakers of American English with no knowledge of tonal languages were recruited. Participants were first trained on Mandarin tone categories (sound-to-category training). Next, they were trained on lexical pseudowords (sound-to-meaning training) over the course of three days. These study procedures are similar to those outlined in Wong and Perrachione (2007). While the lexical task was identical for all participants, each participant was randomly assigned to one of four sound-to-category training paradigms. The four sound-to-category task conditions included the following:

1) in the *Reflect* condition, participants learned to categorize tones with informationally rich (full) feedback and blocked talker presentation aimed to enhance reflective learning

2) in the *Reflex* condition, participants learned to categorize tones with minimally informational (minimal) feedback, and randomized (mixed) talker presentation aimed to enhance reflexive learning

3) in the *Reflect→Reflex* condition, participants learned to categorize tones with full feedback and blocked talker presentation in the initial three blocks of training, and minimal feedback and mixed talker presentation in the final three blocks of training, aimed to initially promote reflective learning and then transfer control to reflexive learning

4) in the *Reflex→Reflect* condition, participants learned to categorize tones with minimal feedback and mixed talker presentation in the first three blocked of training, and full feedback and blocked talker presentation in the final three blocks of training.

Immediately after completing sound-to-category training, participants in all four conditions underwent an identical generalization task where they categorized novel Mandarin tone stimuli produced by novel talkers without receiving feedback. Following this, all participants then underwent five identical sessions of a lexical learning task (sound-to-meaning training). The impact of the four training conditions (Reflect, Reflex, Reflect→Reflex and Reflex→Reflect) was compared for sound-to-category generalization and sound-to-meaning mapping.

Prior work supports that tone categorization is a reflexive-optimal task (Chandrasekaran et al., 2014). Since reflexive learning does not involve working memory, one of the advantages of using this strategy is that it may free up limited cognitive resources that can then be allocated to other things. For instance, if tones are reflexively learned, working memory resources may then be used for word learning. Taking this into account along with the bootstrap interaction theory, the a priori prediction was that initial reflective (full feedback and blocked talkers) and later reflexive-optimal environments (minimal feedback and mixed talkers) experienced by participants in the Reflect→Reflex condition would result in better sound-to-category performance relative to other groups. Participants in the Reflex condition were predicted to perform better than participants in the Reflect condition on the basis that tone categorization tasks are thought to be (ultimately) reflexive-optimal. There is only support for a one-way interaction between the learning systems (Chandrasekaran, Koslov, & Maddox, 2014). Thus, participants in the Reflex→Reflect condition were predicted to not perform as well as participants in the Reflect→Reflex condition, but to perform better than participants who only experienced reflective-optimal training in the Reflect condition. Accurate tone categorization is necessary for selecting appropriate responses in the lexical task where participants map sound-to-meaning. As such, it was predicted that participants in conditions that showed the highest accuracies while categorizing new tone stimuli during the sound-to-category phonological generalization task would also perform better on the sound-to-meaning lexical training tasks.

METHODOLOGY

PHONOLOGICAL CATEGORIZATION TRAINING

Participants

Native speakers of American English (N = 40, 23 females; age range of 18-38 years old, mean age = 21.53; see Table 1) were recruited from the University of Texas at Austin and the greater Austin area. All participants reported no prior history of neurological disorders or hearing problems. All participants passed a pure tone audiological threshold hearing screening < 25 dB HL at 1000, 2000, and 4000 Hz. All participants completed a language history questionnaire and reported no prior exposure to tonal languages (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). A music history questionnaire was collected to exclude potential participants with more than eight years of continuous, formal musical training (adapted from Wong & Perrachione, 2007). Participants were randomly assigned to one of four phonological training conditions detailed in the procedures. Informed consent was obtained from all participants and compensation was provided as approved by the Institutional Review Board at the University of Texas at Austin.

	Reflect→Reflex	**Reflex→Reflect**	**Reflect**	**Reflex**
n	10	10	10	10
Sex (M:F)	5:5	2:8	5:5	5:5
Mean Age (SD)	19.60 (2.50)	21.50 (5.93)	21.10 (2.77)	23.90 (3.07)

Table 1. Participant demographic information for each of the four phonological training conditions.

Stimuli

Stimuli were adapted from a previous study by Chandrasekaran and colleagues (Chandrasekaran, Koslov, & Maddox, 2014). Four native Mandarin Chinese speakers (2 female) produced four Mandarin Chinese lexical tones [tone 1 (T1; high-level), tone 2 (T2; low-rising), tone 3 (T3; low-dipping), and tone 4 (T4; high-falling)]. These tones were produced in citation form in the context of the five monosyllabic words /bu/, /di/, /lu/, /ma/, and /mi/ (Alexander, Wong, & Bradlow, 2005), resulting in a total of eighty unique stimuli. Multiple speakers were used to emulate the inherent variability present in natural language. Stimuli were RMS amplitude (70 dB) and duration (0.44 s) normalized using Praat (Boersma & Weenink, 2005).

Procedure

Participants were provided a copy of the consent form when they arrived. They completed the language and music history questionnaire if they had not done so already. Participants were seated in a sound-attenuated booth and the pure tone hearing screening was conducted. They were then seated in front of a computer in the same sound-attenuated booth for the experimental tasks. All stimuli were presented binaurally through circumaural headphones. Participants were told that they would be listening to sounds that could be categorized into 1, 2, 3, and 4. They were instructed to press a number key on the keyboard that corresponded to the category whenever they heard a sound. Each stimulus was followed by the response prompt, “Which category? (Press the number key)” displayed on the screen. Conditions varied by the type of feedback and talker presentation (see Table 2). Depending on the experiment condition, stimuli were either presented with a blocked (grouped) talker sequence or a mixed (randomized) talker sequence. Additionally, participants received either full or minimal feedback. For example, in conditions with full feedback, participants would either receive “Correct, that was a 2.” or “No, that was a 4.” following their responses. In conditions with minimal feedback, participants would receive either “Correct” or “No” following their responses. Feedback immediately followed participant response and remained on screen for 1s. The stimulus-response-feedback sequence comprised a single trial. One block of phonological training consisted of a randomized presentation of all eighty stimuli. A total of six blocks of eighty trials were presented in the training experiment, yielding four hundred eighty trials for each subject. Participants were given self-timed breaks between each block and selected when to move on to the next block by pressing any button on the keyboard.

Participants were assigned into one of the four conditions: Reflect→Reflex, Reflex→Reflect, Reflect, or Reflex (see Table 2). In the Reflect→Reflex condition (n = 10), stimuli were initially presented with blocked talkers and full feedback for blocks 1-3, targeting the reflective system, followed by mixed talkers and minimal feedback, targeting the reflexive system, for blocks 4-6. In the Reflex→Reflect condition (n = 10), stimuli were initially presented with mixed talkers and minimal feedback for blocks 1-3, targeting the reflexive system, followed by blocked talkers and full feedback, targeting the reflective system, for blocks 4-6. In the Reflect condition (n = 10), stimuli were presented with blocked talkers and full feedback, targeting the reflective system, for all 6 blocks. In the Reflex condition (n = 10), stimuli were presented with mixed talkers and minimal feedback, targeting the reflexive system, for all 6 blocks.

Condition	Early Feedback	Early Talkers	Late Feedback	Late Talkers
Reflect→Reflex	Full	Blocked	Minimal	Mixed
Reflex→Reflect	Minimal	Mixed	Full	Blocked
Reflect	Full	Blocked	Full	Blocked
Reflex	Minimal	Mixed	Minimal	Mixed

Table 2. Feedback type and talker presentation for each condition during early (blocks 1-3) and late (blocks 4-6) phonological categorization training.

Immediately following tone categorization training and prior to the first session of lexical training, participants completed a generalization block where they were presented with novel stimuli that would later be used in lexical training. Participants listened to ninety six pseudowords with superimposed tone stimuli (described in further detail in the next section). They were instructed to categorize the words presented into 1, 2, 3, and 4 based on the tone. The instructions were the same as those provided during phonological categorization training, and participants were instructed to press a number key on the keyboard corresponding to the category after each sound they heard. No feedback was provided during the generalization test.

LEXICAL TRAINING

Participants

All 40 participants from Phase 1 also participated in Phase 2 of the experiment.

Stimuli

Lexical items were selected to simulate how native English speakers learn to use pitch when learning novel words. Stimuli consisted of the English monosyllabic pseudowords /dree/, /fute/, /ner/, /pesh/, /nuck/, and /vece/ with superimposed pitch patterns from Mandarin tones T1, T2, T3, and T4 (Wong & Perrachione, 2007). This resulted in twenty four unique items, which were produced by four native speakers of American English (2 females). Each of the auditory stimuli was associated with one of twenty four images, which referred to a concrete object (e.g., keys, table, scissors). Participants were provided with a sheet that included all twenty four images (adapted from Wong & Perrachione, 2007; see Figure 1). The pitch patterns were produced by a female native speaker of Mandarin Chinese in the context of the word /mi/. Pitch–synchronous overlap and add method was used to superimpose the pitch contour from each of the four Mandarin tones onto the pseudowords (Moulines & Charpentier, 1990). Six pseudowords with the four superimposed pitch patterns (forming six minimal contrast quartets) were created for each speaker, yielding a total of ninety six stimuli. Each stimulus had a duration between 0.35-0.47 s. Stimuli were rated as natural sounding by five native Mandarin Chinese speakers (identification accuracy > 95%).

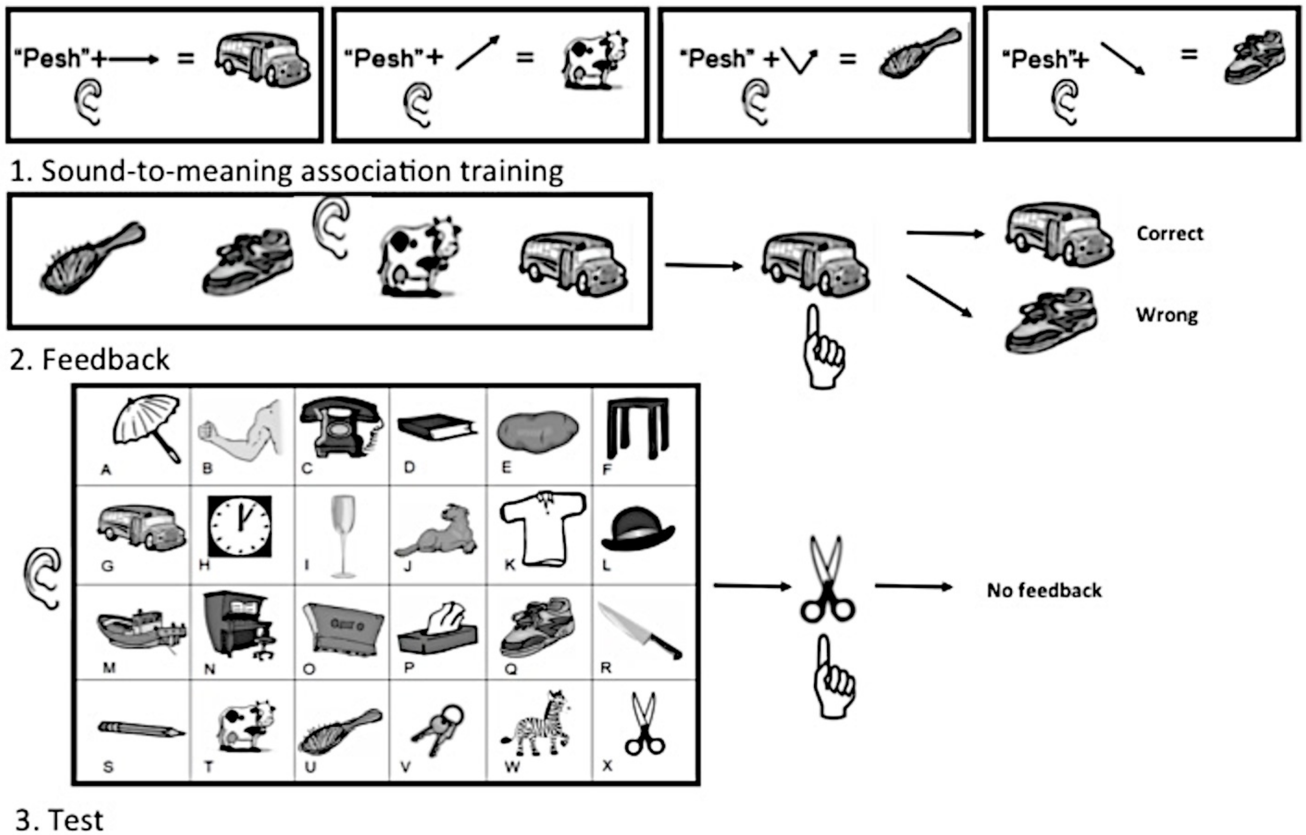

Figure 1. Lexical training task procedure. 1 depicts sound-to-meaning association training, 2 depicts feedback provided on a minimal contrast quartet set, and 3 depicts a trial on the final test block on all twenty four stimuli.

Procedure

All participants completed a total of five sessions of the lexical task over the course of three days. Participants were seated in front of a computer in the same sound-attenuated booth used during phonological training. All stimuli were presented binaurally through circumaural headphones. Participants were instructed that they would be listening to words and that the images that appeared on screen corresponded to these words. They were instructed that they were to select the letter on the keyboard that corresponds to the image and pseudoword + tone stimulus they heard. Stimuli were presented in six blocks, wherein each block consisted of one pseudoword and its four tonal iterations (ex. /pesh/ + T1, /pesh/ + T2, /pesh/ + T3, and /pesh/ + T4), totaling twenty blocks. Training involved the simultaneous presentation of the auditory stimuli with the corresponding image. As each pseudoword + tone combination matched one image on the sheet provided to participants, each block asked participants to respond to novel words for four different objects in the minimal contrast quartet set. Stimuli presentation was randomized for all participants and training block order was randomized across sessions. The experiment was self-paced and feedback followed immediately after participant response for 1 s. After completing all training blocks within one session, participants were presented with a final test block of all ninety six stimuli (all twenty four pseudoword + tone stimuli produced by all four speakers). Participants were again asked to match each auditory stimulus to its corresponding image. Participants were allowed as much time as needed to respond. No feedback was given during the final test block.

EXPERIMENT SCHEDULE

Phonological categorization training and lexical training were conducted over the course of three experiment sessions. Participants signed consent forms, completed music and language history questionnaires, underwent a hearing screening, completed the phonological categorization training session, and completed the first session of the lexical task on day 1. They completed the second and third sessions of the lexical task on day 2. They completed the fourth and fifth sessions of the lexical task on day 3. The three experiment sessions were completed across no more than a four-day period.

RESULTS

PHONOLOGICAL CATEGORIZATION TRAINING

Phonological categorization training performance was assessed across training conditions using mixed effects modeling with a binomial logit link (Bates, Maechler, & Bolker, 2012). The dependent variable was trial-by-trial accuracy, which was coded as either correct or incorrect. The fixed effects included the trial number (one to four hundred eighty) and the four phonological categorization training conditions (Reflect→Reflex, Reflex→Reflect, Reflect, and Reflex,). Based on our predictions, we expected that the Reflect→Reflex group would perform the best overall. The basis behind this prediction was that participants in this condition would first learn to categorize Mandarin tones in an environment optimized for reflective learning, where they could test rules to establish category boundaries, before transitioning in the second half of training to an environment optimized for reflexive learning, which is optimal for non-native speech categorization. The Reflect→Reflex group was set as the reference level. Effects were corrected for random intercepts for each subject in order to correct for the random variance arising from non-systematic individual variability. The estimated coefficient for each simple effect quantified the logit-transformed impact on the likelihood of producing a correct response for a given trial. The results of this analysis are shown in table 3.

Fixed effect	**Estimate**	**Standard error**	***z* value**	***p* value**
(Intercept)	0.4061	0.2938	1.382	0.1669
Trial	0.0031	0.0003	12.138	< 0.0001***
Reflect group	0.0846	0.4156	0.204	0.8387
Reflex group	-0.7803	0.4147	-1.881	0.0599
Reflex→Reflect group	-0.5995	0.4154	-1.443	0.1490
Trial: Reflect group	0.0009	0.0004	2.439	0.0148*
Trial: Reflex group	0.0011	0.0004	3.164	0.0016**
Trial: Reflex→Reflect group	0.0024	0.0004	6.313	< 0.0001***

Table 3. Result of mixed effects modeling for phonological categorization accuracy as a function of training condition and trial. The Reflect→Reflex group was set as the reference. *p<0.05; **p<0.01; ***p<0.001.

The simple effect for trial ($p < .001$) suggests that later trials led to higher accuracies relative to earlier trials for the Reflect→Reflex group. This indicates that participants in this group were able to learn to categorize Mandarin tones more accurately as trials progressed. The Reflect ($p=.0148$), Reflex ($p=.0016$), and Reflex→Reflect ($p < .001$) training groups all showed higher learning rates relative to the Reflect→Reflex group. That is, relative to the Reflect→Reflex group, participants in the other training conditions showed a faster increase in accuracy as trials progressed. Although marginally significant ($p=0.0599$), the Reflex group shows the lowest estimate (-0.7803), which may explain why it appears that this is the lowest performing group. These results are displayed in Figure 2 below.

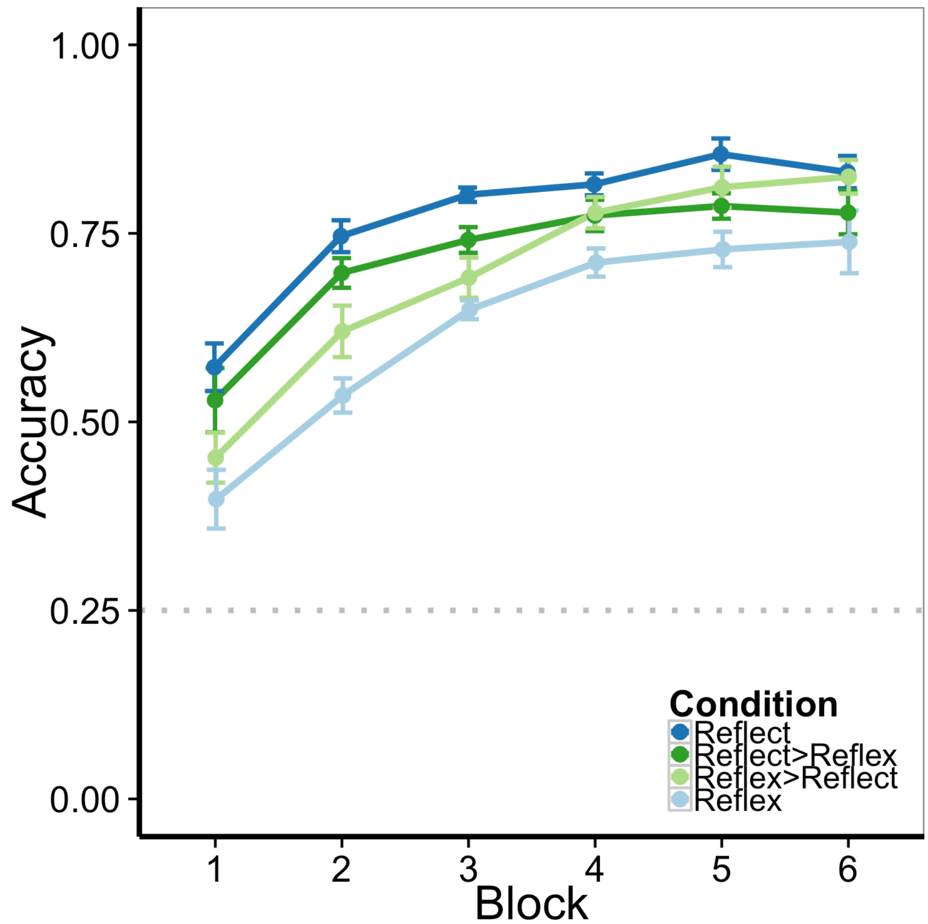

Figure 2. Phonological categorization accuracy for Mandarin tone stimuli across six training blocks. Chance performance (i.e., 25%) is marked by the dotted gray line.

Next, participant's responses on the generalization test where they were presented with each of the novel ninety six pseudoword + tone stimuli with no feedback and asked to categorize the tone were analyzed. At the initial trial, there were no significant differences in tone categorization accuracy found between any of the groups. The simple effect for trial (p=0.448) was not significant, suggesting that later trials did not lead to higher accuracies relative to earlier trials for the Reflect→Reflex group (see Table 4). No interactions for trial and condition were significant: Reflect (p=0.136), Reflex (p=0.110), and Reflex→Reflect (p=0.438). These results suggest that there were no significant differences between groups at either the initial stimuli presentation or with successive trials. Figure 3 shows the generalization test results for all groups.

Fixed effect	**Estimate**	**Standard error**	***z* value**	***p* value**
(Intercept)	0.1106	0.2505	0.442	0.659
Trial	-0.0018	0.0024	-0.758	0.448
Reflect Group	-0.1770	0.3548	-0.499	0.618
Reflex Group	-0.2546	0.3557	-0.716	0.474
Reflex→Reflect Group	0.0065	0.3557	0.018	0.986
Trial: Reflect Group	0.0051	0.0034	1.489	0.136
Trial: Reflex Group	0.0055	0.0035	1.600	0.110
Trial: Reflex→Reflect Group	0.0027	0.0035	0.776	0.438

Table 4. Result of mixed effects modeling for phonological categorization accuracy in the generalization test. The Reflect→Reflex group was set as the reference.

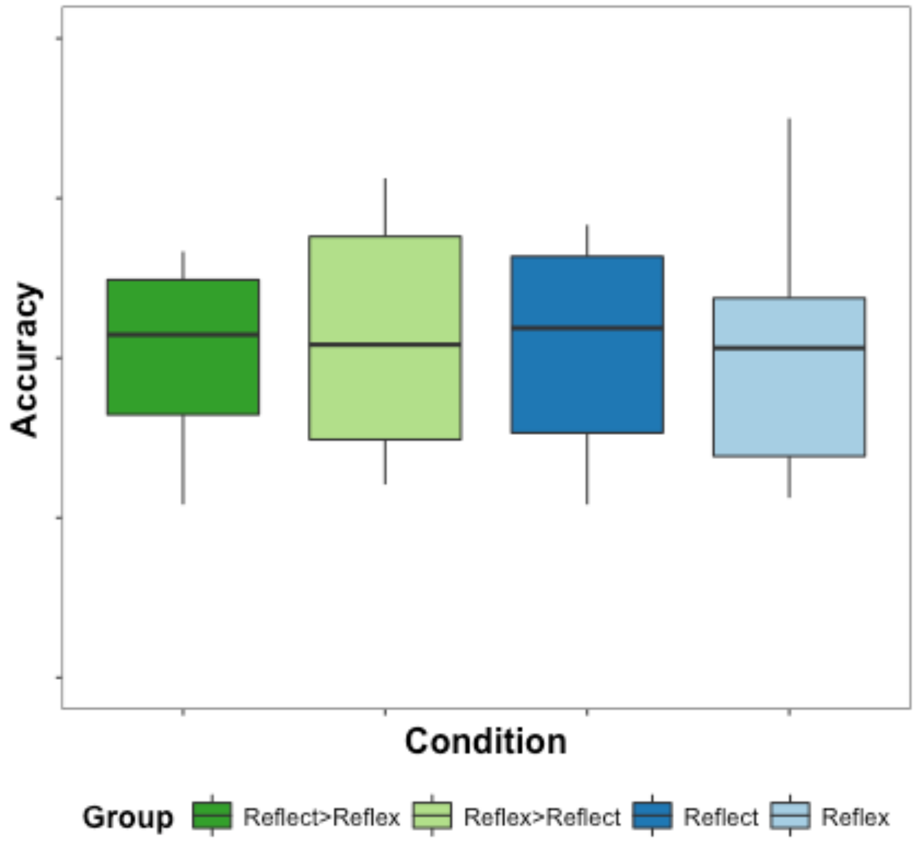

Figure 3. Tone categorization accuracy results for the phonological categorization generalization test.

LEXICAL TRAINING

Performance in the lexical training component of the experiment was compared across groups using mixed effects modeling with binomial logit link (Bates, Maechler, & Bolker, 2012). Session-by-session accuracy served as the dependent variable. Fixed effects included the four training conditions, the session number (1, 2, 3, 4, and 5), and their interaction terms. As with phonological training analyses, the Reflect→Reflex group was set as the reference level. The model was corrected for by-participant random intercept to adjust for non-systematic differences arising from individual variability.

Fixed effect	**Estimate**	**Standard error**	***z* value**	***p* value**
(Intercept)	-3.1931	0.4001	-7.981	< 0.0001***
Session	0.7884	0.0292	26.978	< 0.0001***
Reflect group	0.2469	0.5662	0.436	0.6628
Reflex group	0.2586	0.5652	0.458	0.6482
Reflex→Reflect group	0.4530	0.5642	0.803	0.4221
Session: Reflect group	-0.1655	0.4058	-4.079	< 0.0001***
Session: Reflex group	-0.1619	0.0419	-3.860	0.0001***
Session: Reflex→Reflect group	-0.2108	0.0400	-5.275	< 0.0001***

Table 5. Result of mixed effects modeling for lexical learning task accuracy. The Reflect→Reflex group was set as the reference. *p<0.05; **p<0.01; ***p<0.001.

Participant responses across all five lexical sessions were analyzed. At the initial trial, there were no significant differences found between any of the groups. The significant effect of session number suggests an increased likelihood of producing an accurate response as participants progress through lexical training sessions (p<0.0001). This signifies that participants in the Reflect→Reflex group were better able to associate the stimuli with the correct pseudoword + tone combination with continued lexical training sessions. All interactions for session and condition were significant, Reflect (p< 0.0001), Reflex (p=0.0001), and Reflex→Reflect (p< 0.0001); however, the estimates are all negative. This suggests that the learning rates for these conditions were lower than the learning rate for individuals in the Reflect→Reflex condition. The magnitude of these interaction effects suggests that while the participants in all four groups performed better on the novel lexical items across trials, the Reflect→Reflex group displayed the greatest increase in accuracy as the sessions progressed. The Reflex→Reflect group was observed to show the lowest increase across sessions ($p < 0.001$). These results are displayed in Figure 4 below.

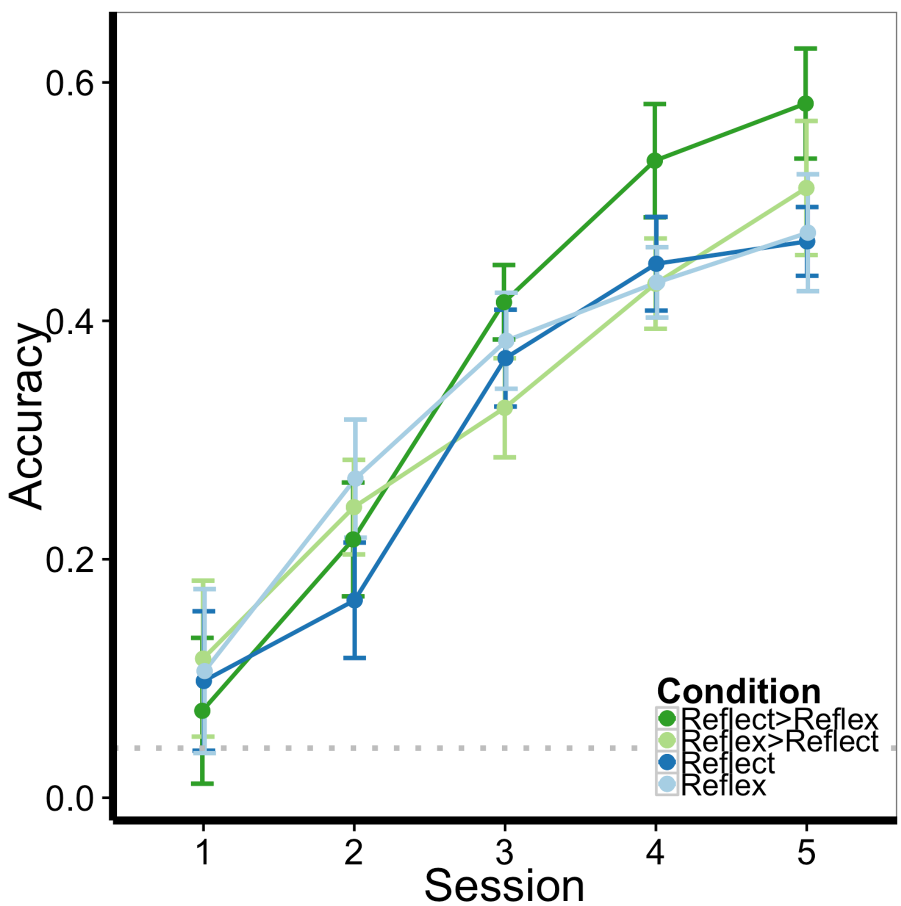

Figure 4. Lexical training accuracy for pseudoword + tone stimuli for the four phonological training conditions. Error bars denote standard error.

Tone and Word Error Analyses

The results described above take into account total accuracy on the lexical task, wherein both the correct pseudoword and the correct Mandarin tone category must be selected in order for the participant's response to be counted as correct. Due to the fact that either the pseudoword or the tone category may be correct for a particular item, additional error analyses were examined for tones and pseudowords individually. The goal of these analyses was break down the nature of heterogeneity in the performance profile across groups in order to determine whether one aspect of the stimuli (tone or pseudoword identification) could explain the patterns seen in lexical training accuracy. Two separate analyses were run. First, tone errors were examined. These are responses that were accurate for the pseudoword but inaccurate for the tone. In a separate analysis, word errors were examined. These are responses that were accurate for the tone but inaccurate for the pseudoword.

As in the lexical training analysis, mixed effects modeling with binomial logit link with the same specified parameters was used to compare tone errors across groups. The Reflect→Reflex group was set as the reference level. The simple effect of session was positive and significant ($p<0.0001$), indicating that the proportion of tone errors increased across sessions for participants in this group (see Table 6). No one group showed significantly less or more tone errors as sessions progressed: Reflect ($p=0.285$), Reflex ($p=0.139$), and Reflex→Reflect ($p=0.241$).

Fixed effect	**Estimate**	**Standard error**	***z* value**	***p* value**
(Intercept)	-1.3263	0.2812	-4.716	< 0.0001***
Session	0.0962	0.0235	4.056	< 0.0001***
Reflect group	-0.5454	0.3997	-1.365	0.172
Reflex group	-0.5444	0.4008	-1.358	0.174
Reflex→Reflect group	-0.6579	0.4013	-1.639	0.101
Session: Reflect group	0.0377	0.3523	1.069	0.285
Session: Reflex group	0.0524	0.0354	1.480	0.139
Session: Reflex→Reflect group	-0.0432	0.0369	-1.172	0.241

Table 6. Result of mixed effects modeling for tone errors in the lexical learning task.

*p<0.05; **p<0.01; ***p<0.001.

An analysis was run using the same mixed effects modeling with binomial logit link this time examining word errors. These results are shown in Table 7. The simple effect of session was found to be positive and significant (p<0.0001). The negative estimate for session indicates that as participants in the Reflect→Reflex group progress through lexical learning sessions, they make a smaller proportion of word errors. Unlike with tone errors, there were significant group differences found in the word error analysis. Relative to participants in the Reflect→Reflex group, word errors decreased less across sessions for all other groups: Reflect (p< 0.0001), Reflex (p< 0.0001), and Reflex→Reflect (p< 0.0001).

Fixed effect	**Estimate**	**Standard error**	***z* value**	***p* value**
(Intercept)	-0.6367	0.2514	-2.532	0.011*
Session	-0.5058	0.0349	-14.480	< 0.0001***
Reflect group	-0.0774	0.3529	-0.219	0.8263
Reflex group	-0.4212	0.3552	-1.186	0.2356
Reflex→Reflect group	0.3816	0.3526	1.082	0.2790
Session: Reflect group	0.2472	0.0441	5.611	< 0.0001***
Session: Reflex group	0.2573	0.0454	4.542	< 0.0001***
Session: Reflex→Reflect group	0.2573	0.0454	5.669	< 0.0001***

Table 7. Result of mixed effects modeling for word errors in the lexical learning task. The Reflect→Reflex group is set as the reference. *p<0.05; **p<0.01; ***p<0.001.

Taken together, these results suggest that across the multiple lexical training sessions, participants in the Reflect→Reflex group were better able to learn to simultaneously identify pseudowords and tones for a given stimulus, relative to those in the Reflect, Reflex, and Reflex→Reflect conditions. This may be driven by results of the word error analysis, showing that participants in the Reflect→Reflex group made significantly fewer pseudoword identification errors as they progressed through lexical training sessions.

DISCUSSION

The findings of this experiment show that we can optimize lexical learning by manipulating the phonological training environment. This study focused on manipulating the categorization training environment for non-native Mandarin tones, to optimize learning outcomes for novel pseudowords that utilize these contrasts. The first variable manipulated in phonological training paradigms was the amount of information given in feedback (full v. minimal feedback). Full feedback is thought to support reflective learning, while minimal feedback is thought to support reflexive learning (Maddox, Love, Glass, & Filoteo, 2008; Chandrasekaran, Yi, & Maddox, 2014). The second variable manipulated in phonological training paradigms was the order of talker presentation (blocked v. mixed talkers; Chandrasekaran, Yi, & Maddox, 2014). Blocked talkers have been linked to reflective learning, while mixed talkers have been linked to reflexive learning. Individuals were randomly assigned to four phonological training paradigms. Two conditions had a consistent environment throughout all blocks of training: individuals in the Reflect condition received full feedback and blocked talkers; individuals in the Reflex condition received minimal feedback and mixed talkers. The other two conditions had environments that switched halfway through training after three blocks. Participants in the *Reflect→Reflex* condition initially received full feedback and blocked talkers and later received minimal feedback and mixed talkers. Participants in the Reflex→Reflect condition initially received minimal feedback and minimal talkers and later received full feedback and blocked talkers. It is important to note that participants in all conditions underwent an identical phonological generalization task and five identical sessions of lexical training.

All groups performed above chance on the phonological training task. This means that participants in all groups, regardless of feedback type and talker presentation provided, were able to learn to categorize Mandarin tones above chance. Also, participants in all groups provided more accurate responses in later trials relative to earlier trials. Against predictions, participants in the Reflect, Reflex, and Reflex→Reflect conditions showed higher learning rates relative to participants in the Reflect→Reflex group. Participants in these groups were able to increase accurate categorization responses more quickly as trials progressed. Prior research has shown that participants in a reflex-optimal environment perform best on tone categorization tasks. This group did not have the best performance; however, this group had the lowest estimate before the initial trial, meaning that prior to training, performance in this group was estimated to be lower than others, which may in part explain why performance for this group was lower than expected. Following the sound-to-category training, participants in all groups completed the generalization task with the stimuli that would later be used in the lexical training tasks. There were no significant differences in the categorization accuracy between any of the groups. This suggests that 1) all groups were able to successfully categorize Mandarin tones above chance accuracy and 2) all groups had relatively similar (and stable) accuracy categorizing novel stimuli. The lack of between condition differences in the generalization task suggests that while different training paradigms may be more or less beneficial during tone categorization training, paradigms that target reflective learning, reflexive learning, or both types of learning are able to result in successful non-native speech sound categorization.

While participants in all conditions had similar performance on categorizing novel stimuli, there were significant performance differences between conditions on the lexical task. Participants in all four conditions had improved accuracy on lexical items as sessions progressed. Thus, regardless of what type of feedback and talker presentation participants received, participants were able to learn to categorize Mandarin tones. Although there were no significant group differences prior to lexical training, participants in the Reflect→Reflex group learned to associate the pseudoword + tone combination with the correct image the most successfully with continued lexical training sessions. The Reflex→Reflect group showed the lowest accuracy increase across successive sessions. The learning rates for all other conditions were lower than that of the Reflect→Reflex condition.

In addition to having the highest accuracy across the greatest number of lexical training sessions and having the highest learning rate for lexical items, participants in the Reflect→Reflex group also made significantly fewer word errors relative to other groups. While the proportion of word errors decreased across lexical training sessions for all groups, participants in the Reflect→Reflex group made significantly fewer word errors relative to other groups. As participants in all conditions completed more lexical sessions, they showed an increase in tone errors. Consistent with performance in the generalization task, no condition significantly differed from one another in tone errors. Participants in the Reflect→Reflex group may have higher accuracy on the novel pseudoword + tone items throughout training due to fewer pseudoword identification errors in succeeding lexical training sessions. Participants are able to solely focus on correctly categorizing the Mandarin tones during the phonological categorization training and generalization test. In the lexical task, participants must not only successfully select the correct phonological category but also the correct pseudoword. Perhaps more tone errors seen in lexical training tasks are due to a tradeoff, as participants must simultaneously attend to selecting the correct tone category as well as selecting the correct pseudoword.

Notably, while the Reflect→Reflex group showed the highest accuracy in the lexical training tasks, this group did not attain the highest accuracy during sound-to-category phonological training and did not differ from the other three groups on the phonological generalization task. This suggests that while a given training environment may not be immediately beneficial, the environment may still be associated with improved learning later on. Based on the bootstrap interaction theory, early reflective learning is expected to bootstrap and enhance reflexive-optimal task performance (Paul & Ashby, 2013). Results from the phonological training portion of our study do not support this. Participants in the Reflect→Reflex group who were exposed to a training environment that supported reflective learning and later reflexive learning for a task that has been found to be reflexive-optimal did not have the best performance on tone categorization in training. Additionally, they had equivalent performance to all other groups in a generalization task. Participants in the Reflect→Reflex group had a training paradigm that initially supported reflective and later reflexive learning performed best on the lexical training. Participants in the Reflex→Reflect group, who had the opposite order of training environments, showed the lowest increase in accuracy across successive lexical sessions. An important advantage of learning using the reflexive system is that working memory is not involved, as it is in the reflective system. Participants who learn tones reflexively are then able to allocate more of the limited working memory resources towards learning the novel words in the lexical task. When learners are able to automatize motor behaviors cortical resources for other cognitive processes may be freed (Chandrasekaran et al., 2015). To succeed in word learning, learners must be able to distinguish categories and map sounds onto objects. In our study, participants in all conditions were able to categorize tones for novel stimuli with similar success. If, however, this categorization is less resource intensive, then more working memory resources can be allocated to word learning, which can lead to improved performance. This advantage, combined with the one-way interaction between learning systems posited by on the bootstrap interaction theory, support that it may be the case that initially using reflective strategies and transitioning to reflexive strategies later in training may leverage interactions between the systems for better learning (Chandrasekaran, Koslov, & Maddox, 2014).

LIMITATIONS AND FUTURE DIRECTIONS

There are several limitations to the current study and future studies that could be conducted to address these. First, the sample size was limited, with only forty participants completing phonological and lexical training tasks. A control condition with no sound-to-category phonological training was not included. Future studies should consider including a no training condition, as it cannot be determined whether participants would perform above chance level on the lexical task without prior exposure to Mandarin tone stimuli or prior sound-to-category training. Results from this study show that regardless of what training paradigm a participant received, all participants were able to learn to categorize Mandarin tones successfully. Training paradigms employed here targeted only one system at a time by presenting a talker presentation and feedback type that have previously been found to promote the use of one learning system. It would be interesting to examine how performance differs when these manipulations are mixed across learning systems (i.e., comparing conditions presenting mixed talkers with full feedback with conditions presenting blocked talkers with minimal feedback) versus when they are consistent for the learning systems (as in this study). Future studies could also incorporate retention tasks after the initial training period. This could provide insight into how well these non-native speech sounds are learned, and whether one training paradigm may lead to better tone categorization retention and/or better lexical performance once time has passed. Potential studies could also replicate this design, including the carryover to lexical items, using different non-native speech sounds, such as using the /r/ and /l/ contrast for Japanese speakers or using the Hindi dental and retroflex contrast for English speakers (Pruitt, Jenkins, & Strange, 2006; Lively et al., 1993; Lively et al., 1994, McClelland et al., 2002).

CONCLUSION

The study provides evidence that a particular phonological training environment can enhance lexical learning. Participants who first learn to categorize tones in a reflectively and later reflexively learn novel words more successfully. The presumed process is that learning to categorize tones reflexively may allow for greater allocation of working memory resources that can then be utilized in word learning. This advantage, combined with the one-way interaction between the reflective and reflexive system, can leverage interactions between the systems to support optimal learning. Participants who learn to categorize tones in a training environment that early on supports reflective learning by providing full feedback and blocked talkers may allow participants to test rules and begin forming category boundaries. By changing the training environment for later learning to support reflexive learning and providing minimal feedback and mixed talkers, learners are encouraged to use the reflexive system, which is thought to be optimal for the tone categorization task. While manipulation of the training environment to selectively target reflective and/or reflexive learning systems did not lead to better performance during phonological categorization, it did lead to greater success learning lexical items. This suggests something beyond speech sound categorization accuracy, perhaps the ability to switch between learning strategies or more working memory resource availability, may be occurring that affects how well lexical items are learned. Since the lexical training paradigm was identical for all participants, performance differences can be attributed to the preceding phonological training paradigm. The results of this study suggest that reducing challenges by manipulating the phonological training environment may be able to bootstrap lexical learning.

References

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *Proceedings of Interspeech 2005-Eurospeech-9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 397–400.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204-210.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147-161.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Netherlands: Springer.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171-204.

Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Attention, Perception, & Psychophysics*, *29*(3), 191-211.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, 1334.

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (version 4.3. 14). [Computer program].

Bradlow, A. R. (2008). Training non-native language sound patterns. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, pp. 287–308. Amsterdam: Benjamins.

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 & Psychophysics, 61,* 977– 985.

Chandrasekaran, B., Koslov, S., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in Psychology, 5*, 825.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*, 148-156.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of Acoustical Society of America, 128*(1), 456-465.

Chandrasekaran, B., Yi, H. G., Blanco, N., McGeary, J. E., & Maddox, W. T. (2015). Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2. *Journal of Neuroscience, 35*(20), 7808-7812.

Chandrasekaran, B., Yi, H., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review, 21*(2), 488-495.

Chao, Y. R. 1968. *A grammar of spoken Chinese*, Berkeley, CA: University of California Press.

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.

Ellis, N. C. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(1), 234-250.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78-104.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268-294.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349.

Gandour, J. T. (1978). The perception of tone. *Tone: A linguistic survey*, 41-76.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1-33.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575-582.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099-3111.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*(1), 42-46.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics, 72*(5), 1218-1227.

Hume, E., & Johnson, K. (2001). A model of the interplay of speech perception and phonology. In E. Hume & K. Johnson (eds.), *The role of speech perception in phonology* (pp. 3– 26). San Diego, CA: Academic Press.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research, 56*(3), 1045-1050.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215-258.

Jongman, A., & Moore, C. B. (2000). The role of language experience in speaker and rate normalization processes. In I*NTERSPEECH* (pp. 62-65).

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2-4), 63-67.

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences*, *3*(4), 123-124.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161-168.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*(2), 263-285.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: John Wiley.

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science, 35,* 1390–1405.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*(1), 3-11.

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. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *The Journal of the acoustical society of America*, *96*(4), 2076-2087.

Maddox, W. T. & Chandrasekaran, B. (2014). Tests of a dual-systems model of speech category learning. *Bilingualism: Language and Cognition, 17*(4), 709-728.

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging, 28*(4), 1042.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, *108*(2), 578-589.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967.

Massaro, D. W., Cohen, M. M., & Tseng, C. Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 267-289.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior, 77*(4-5), 657-662.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*(5-6), 453-467.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... & Allik, J. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432.

Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 279-291.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331-347.

Paul, E. J., & Ashby, F. G. (2015). A neurocomputational theory of how explicit learning bootstraps early procedural learning. *Frontiers in Computational Neuroscience, 7*(177, 1-17).

Penfield, W. & Roberts, L. (1959). *Speech and brain mechanisms*. New York: Atheneum Press.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (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*(1), 461-472.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684-1696.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203-219.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892-1902.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, *18*(6), 1029-1043.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences, 104*(33), 13273-13278.

Ventura-Campos, N., Sanjuán, A., González, J., Palomar-García, M. Á., Rodríguez-Pujadas, A., Sebastián-Gallés, N., ... & Ávila, C. (2013). Spontaneous brain activity predicts learning ability of foreign sounds. *Journal of Neuroscience*, *33*(22), 9295-9305.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168-176.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173-196.

Wong, F. C., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *Journal of Neuroscience*, *31*(24), 8780-8785.

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

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.

Yi, H. G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2014). The role of corticostriatal systems in speech category learning. *Cerebral Cortex, 26* (4), 1409-1420.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage, 46,* 226– 240.