LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Pitch Ability As an Aptitude for Tone Learning

Anita R. Bowles,[a] Charles B. Chang,[b] and Valerie P. Karuzis[a]

[a]University of Maryland and [b]Boston University

Tone languages such as Mandarin use voice pitch to signal lexical contrasts, presenting a challenge for second/foreign language (L2) learners whose native languages do not use pitch in this manner. The present study examined components of an aptitude for mastering L2 lexical tone. Native English speakers with no previous tone language experience completed a Mandarin word learning task, as well as tests of pitch ability, musicality, L2 aptitude, and general cognitive ability. Pitch ability measures improved predictions of learning performance beyond musicality, L2 aptitude, and general cognitive ability and also predicted transfer of learning to new talkers. In sum, although certain nontonal measures help predict successful tone learning, the central components of tonal aptitude are pitch-specific perceptual measures.

Keywords second language; aptitude; tone learning; pitch perception; Mandarin Chinese; musical experience

Introduction

For adults, mastering a second/foreign language (L2) can be challenging. Whereas some reach high levels of competence in one or more L2s,

---

Anita Bowles is now at Rosetta Stone. The findings reported in this article are based on work supported, in whole or in part, by funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are our own and do not necessarily reflect the views of any of our institutions or any agency of the United States Government. We thank Jessica Bauman, Alia Biller, Allison Blodgett, Greg Colflesh, Ryan Corbett, Joseph Dien, Dimitrios Donavos, Meredith Hughes, Anton Rytting, and Matt Winn for their work on the larger research project associated with this study and Price Bingham, Janet Cook, Eli Cooper, Meg Eden, Christian Frable, and Lora Grasso for help with data collection. Finally, we thank Henk Haarmann, four anonymous reviewers, and other readers for support and comments on earlier versions of this manuscript and Bob Slevc for advice on musical aptitude and for sharing Wing Musical Aptitude Test stimuli.

Correspondence concerning this article should be addressed to Anita Bowles, Rosetta Stone, 135 W. Market Street, Harrisonburg, VA 22801. E-mail: anita.r.bowles@gmail.com

Language Learning 66:4, December 2016, pp. 774–808

DOI: 10.1111/lang.12159

others remain stuck at beginner or intermediate levels, particularly in aspects of phonology and morphosyntax (Abrahamsson & Hyltenstam, 2008; Harley & Wang, 1997; Johnson & Newport, 1989). During the 20th century, a number of tests were developed to predict which adult learners would be most successful at L2 learning. These include the Army Language Aptitude Test (Carroll, 1981; Neumann, Abrahams, & Githens, 1968) and its successor, the Defense Language Aptitude Battery (DLAB; Petersen & Al-Haik, 1976), the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959), the Pimsleur Language Aptitude Battery (Pimsleur, 1966), and the VORD[1] (Parry & Child, 1990). These tests were designed to measure the construct of foreign language aptitude, which Carroll (1981) described as "the individual's initial state of readiness and capacity for learning a foreign language, and probable degree of facility in doing so" (p. 86). Thus, L2 aptitude encompasses two related, but separable, dimensions: capacity, which defines the level of eventual mastery that the learner can achieve (given optimal learning conditions), and facility, which refers to how quickly the learner will progress.

Existing aptitude tests have primarily focused on measuring rate of attainment (i.e., the facility dimension), often for practical reasons. For example, tests developed by the U.S. Army (e.g., DLAB) are used to inform the selection of military personnel for assignment to language training, based on their likelihood of successfully reaching a criterion level of proficiency within a given number of weeks (Carroll, 1981; Petersen & Al-Haik, 1976). Recently, research into the capacity dimension has led to the development of the High-Level Language Aptitude Battery (Hi-LAB), a test designed to measure a learner's ability to reach nativelike proficiency in a L2 (Doughty et al., 2010; Linck et al., 2013). Development of the Hi-LAB took place during a resurgence in research on L2 aptitude. Over the past two decades, work in this area has accelerated. In addition to research on the capacity dimension of L2 aptitude (e.g., Abrahamsson & Hyltenstam, 2008; Linck et al., 2013; Robinson, 2005), researchers have reconsidered the theoretical bases of L2 aptitude to update the construct, using findings from cognitive psychology and L2 acquisition (Miyake & Friedman, 1998; Robinson, 2001; Skehan, 2002; Sparks, Patton, Ganschow, & Humbach, 2011).

Notably, existing aptitude tests were designed to measure general L2 aptitude (i.e., one's overall aptitude for learning any L2, given sufficient motivation and appropriate instruction), not aptitude for specific languages. In the present study, we explored the possibility of refining measures of language aptitude by focusing on specific features of the target language known to be particularly challenging to L2 learners. In the next section, we explain the rationale behind this approach in further detail and apply the approach to L2 learning of lexical tone.

Background

Differential Prediction of L2 Aptitude

Although available aptitude tests have been validated with adult learners of diverse L2s, to date no test can determine whether a particular student's aptitude may vary for different L2s. For example, might one student do well in Arabic but struggle with Korean, or vice versa? This type of *differential prediction* is challenging, because languages vary along many dimensions including their typological distance from the native language (L1) and the relative complexity of their orthographic, phonological, and syntactic systems (see MacWhinney, 1996, for a review of these issues).

The DLAB was originally intended to provide language-specific predictive validity, but ultimately no test (including the DLAB) provides differential prediction by language. Currently, some attempt at matching students to language is made by the U.S. military, which uses DLAB scores as one criterion for assigning students to languages in four categories of difficulty for native English speakers (Lett & O'Mara, 1990). The easiest languages (Category I) are generally Indo-European languages with a familiar orthography and a fair number of English cognates, whereas the most difficult languages (Category IV) are typologically distant from English and employ a non-Roman script. Matching students to language categories using this strategy appears to be successful (Lett & O'Mara, 1990). However, the question remains whether a more efficient system might be developed to match learners' diverse profiles of strengths and weaknesses to specific languages. Periodic attempts have been made to identify aptitudes important for specific languages (Asher, 1972; Lett & O'Mara, 1990) or typologically similar groups of languages (Neumann et al., 1968), but they have not yielded clear results.

Identifying the set of aptitudes relevant to specific languages is difficult for several reasons, including the number of learners (observations) required and the need to control for language experience and curriculum across languages. However, we suggest that the primary reason for failure to identify aptitudes for specific languages is the granularity of the question. Any given language consists of many typological features, each of which may be challenging to L2 learners for various reasons. Thus, the proper level of analysis is not an entire L2, but the set of linguistic features that constitute an individual language, such as subtle phonological contrasts, a new orthography, a complex morphological system, or obligatory pragmatic features (MacWhinney, 1996). Thus, a more promising (and tractable) research question is whether aptitudes can be identified for linguistic features. If so, it follows that a given L2 could be evaluated in terms of its features and the alignment between those features and students' aptitudes, to inform selection into a language and/or instructional interventions. As a test case, the current study examined aptitude for a crosslinguistically common feature (found in over 40% of the world's languages; Maddieson, 2013)—namely, lexical tone.

Lexical Tone

Unlike stress and intonation languages such as English (where voice pitch serves primarily as a cue for syllabic prominence and/or sentence type), tone languages such as Mandarin employ pitch to signal lexical contrasts (i.e., different words). Much like segment inventories, tone inventories vary considerably across languages, both in terms of the number of categories and the nature of the distinctions among categories (e.g., relative importance of pitch height vs. pitch contour). Two major types of lexical tone systems are register systems, in which tone categories differ primarily in terms of the level (i.e., height) of the pitch contour, and contour systems, in which tone categories differ primarily in terms of the shape of the pitch contour. In addition, some languages present mixed tone systems that include contrasts in pitch level and pitch shape. For example, whereas Yoruba evinces a prototypical register tone system that distinguishes among three phonemic level tones (low, mid, high), Thai has a mixed tone system that distinguishes among five phonemic tones: three level tones (low, mid, high) plus two contour tones (rising, falling) (Maddieson, 2013).

The four-tone system of Mandarin (a contour tone system with high level, mid rising, low falling-rising, and high falling tones) was selected for the current study for several reasons. First, Mandarin is a L2 of increasing importance worldwide (Dillon, 2010; Starr, 2009), so findings on Mandarin are relevant for many language learners. Second, the Mandarin tone system is widely studied in the literature on tone learning; consequently, there is ample basis for comparison with previous results. Third, previous work on tone learning suggests that a contour tone system as in Mandarin may be particularly challenging for English speakers (the target population in this study) because at least older English speakers tend to be biased toward attending to pitch height rather than pitch direction (e.g., Maddox, Chandrasekaran, Smayda, & Yi, 2013). Finally, the Mandarin system is intermediate in size and complexity compared to other tone systems across the world, thus providing a useful starting point for an examination of tonal aptitude.

Indicators of Tonal Aptitude

Although some research has examined the prediction of success in learning Mandarin (Carroll, 1958; Carroll & Sapon, 1955; Winke, 2007, 2013), these studies measured overall proficiency or broad language skills (reading, listening, speaking, writing), rather than success with tone specifically. Thus, existing findings support the predictive validity of the MLAT for Mandarin learners but do not provide information about tonal aptitude.

Because tone languages employ distinctions in relative pitch, one potential measure of tonal aptitude might be ability to perceive relative pitch in either linguistic or nonlinguistic (e.g., musical) contexts. Several strands of evidence suggest a close relationship between musical tone processing and linguistic tone processing. For example, perfect pitch is more prevalent among speakers of tone languages than speakers of nontonal languages (Deutsch, Henthorn, & Dolson, 2004). Furthermore, the strength of the human brainstem's responses to both musical and linguistic pitch are stronger for musicians than nonmusicians (Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007), and early musical training affects later pitch processing in both music and language (Bidelman, Hutka, & Moreno, 2013; Schön, Magne, & Besson, 2004).

Several studies have investigated the relationship of musical experience to lexical tone learning. Results suggest that English-speaking musicians are better than nonmusicians at identifying, discriminating, and imitating Mandarin tones (Alexander, Wong, & Bradlow, 2005; Bidelman et al., 2013; Gottfried & Ouyang, 2005, 2006; Gottfried, Staby, & Ziemer, 2004) as well as identifying Thai tones (Wayland, Herrera, & Kaan, 2010). In addition, English-speaking musicians outperform nonmusicians in a Mandarin word learning task (Wong & Perrachione, 2007; see Cooper & Wang, 2012, for similar results for Cantonese), a disparity associated with differences in the neural processing of pitch (Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong et al., 2007). In these studies of tonal word learning, naïve speakers of English learned a small set of monosyllabic pseudowords (15–24 items total), subsets of which could be distinguished only by tone.

Whereas most recent work on the relationship between music and lexical tone learning has focused on the role of musical experience, other research has examined the relationship of musical aptitude to L2 learning.[2] However, these studies have focused primarily on nontonal languages such as French (Dexter, 1934; Fenner, 1955; Leutenegger & Mueller, 1964; Pimsleur, Stockwell, & Cromley, 1962), English (Brutten, Angelis, & Perkins, 1985; Slevc & Miyake, 2006), Spanish, and Korean (Gilleece, 2006) and have produced a complex set of findings, perhaps due to variability in participant samples and measures of aptitude.[3] For example, Brutten et al. (1985) found that rhythm discrimination, but not pitch discrimination, was correlated with success in English pronunciation (for learners from various language backgrounds), whereas Slevc and Miyake (2006) showed that high overall musical aptitude was associated with successful English /l/-/r/ pronunciation for Japanese learners. Pimsleur et al. (1962) found that accurate pitch discrimination was associated with successful auditory comprehension in French and with high scores in language lab activities, but not with oral proficiency. Most relevant to the current study, both musical experience and aptitude have been shown to predict success with tone: Wong and Perrachione (2007) reported that individuals with greater amounts of private music lessons performed more successfully in a Mandarin tone learning task, and Cooper and Wang (2012) found that both measures of prior musical experience and musical aptitude scores predicted success in tonal word learning for native speakers of English. Additionally, Delogu, Lampis, and Belardinelli (2006, 2010) showed that higher musical aptitude scores were associated with greater tone discrimination ability.

Another potential component of tonal aptitude is the tonal aspect of auditory working memory (WM; e.g., Tierney & Pisoni, 2003). Tanaka and Nakamura (2004) suggest that verbal memory and musical memory are elements of a unitary auditory memory system and that differences in the capacity of this system, as measured by musical and verbal tasks, may predict individual differences in L2 pronunciation. Similarly, Pechmann and Mohr (1992) propose that a tonal loop may support rehearsal of tonal information within the WM system, similar to the way that the articulatory loop supports verbal rehearsal and the visuo-spatial sketchpad supports rehearsal of visual information in Baddeley's WM model (Baddeley, 1986). Additional evidence suggests that musicians, who show enhanced ability to discriminate lexical tones, outperform nonmusicians in behavioral tests of both pitch short-term memory (STM) and some aspects of WM (Bidelman et al., 2013). Other researchers, however, have not found a strong relationship between WM and learning of tonal vocabulary (Perrachione, Lee, Ha, & Wong, 2011).

Taken together, linguistic and musical aspects of pitch ability may relate to tone learning success in four main ways. First, linguistic tone sensitivity and musical pitch sensitivity may reflect a global neural sensitivity to pitch (shared across language and music), which stems from a domain-general cognitive processing of pitch and thus aids in the learning of pitch patterns in general (e.g., Perrachione, Fedorenko, Vinke, Gibson, & Dilley, 2013). Second, linguistic tone sensitivity in particular is likely to provide an advantage in associating linguistic tones with meaning. Third, musical experience may lead to the improvement of pitch processing and/or auditory memory, which in turn may facilitate tone learning (e.g., Patel, 2011). Finally, the pitch aspects of musical aptitude, as a construct separate from previous musical experience, might represent a facet of pitch ability distinct from linguistic tone sensitivity, which (like linguistic tone sensitivity) also drives differences in tone learning success.

The Present Study

Although much of the research on tone perception and learning has focused on group-level contrasts, individual differences in pitch ability are likely to inform the prediction of individual success in learning tones. Such individual differences have been measured behaviorally using various methods, such as tests of pitch perception, pitch discrimination, and pitch STM (e.g., Bidelman et al., 2013; Perrachione et al., 2011; Wong & Perrachione, 2007); however, the literature has not clearly separated linguistic and nonlinguistic (namely, musical) aspects of pitch ability—from each other or, for that matter, from more general cognitive ability—to determine which aspects of pitch ability may be most important for learning a tone language. This limitation of the literature provided the motivation for the current study, which addressed three research questions:

1. Can pitch ability and/or musicality predict tonal word learning over and above general cognitive ability[4] and general L2 aptitude?
2. Which aspects of pitch ability are the most effective predictors of tonal word learning?
3. What is the relationship among pitch ability, musicality, general L2 aptitude, and general cognitive ability?

To answer the first and the second questions, we conducted a correlational study of Mandarin word learning. In this study, naïve native speakers of English completed a battery of cognitive tests and learned a small lexicon of Mandarin pseudowords over six training sessions. Although the basic word learning paradigm was similar to that used in previous work (Chandrasekaran, Sampath, & Wong, 2010; Wong & Perrachione, 2007; Wong et al., 2011), this study moved beyond previous studies in three main ways. First, it employed multiple regression with a sample of learners large enough to treat previously studied predictor variables such as musical experience, as well as learning outcomes, at an individual, rather than group, level (similar to the type of individual-level analysis seen in Perrachione et al., 2011). Second, it jointly examined measures of musical experience, musical aptitude, pitch perception, auditory memory, general cognitive ability, and general L2 aptitude. Third, it used target lexical items that were longer, more phonologically diverse, and more ecologically valid (being consistent with Mandarin phonotactics) than those examined in previous work. These design features allowed us to carry out more fine-grained analyses of individual differences in tonal word learning, to evaluate the predictive power of pitch ability in the context of other relevant predictors, to determine which aspects of pitch ability best predict success with lexical tone, and to arrive at results more generalizable to Mandarin learning outside the confines of our word learning task (e.g., the learning of real Mandarin words).

To address the second and third questions, we conducted a principal component analysis (PCA) targeting the measures collected through the current test battery. This analysis provided information about the interrelationships among included measures, with implications both for scientific knowledge of the connections among pitch perception, pitch memory, and musical measures and for the development of language aptitude tests.

Our regression and PCAs were designed to test three main hypotheses. First, given that general cognitive ability and general L2 aptitude should both predict success at a Mandarin word learning task (as these abilities are important for any L2 learning task), we hypothesized that measures of linguistic and nonlinguistic pitch processing would provide additional predictive validity because of the close relationship of pitch ability to the tonal aspects of the learning task. Second, we hypothesized that pitch-specific behavioral measures (e.g., tone discrimination) would be more effective predictors, compared to non-pitch-specific musical aptitude measures or musical experience measures, because they more directly reflect pitch processing capabilities. Third, we hypothesized that transfer of learning to new talkers (speaking voices) would require additional skills (e.g., ability to classify the physical signal into broad categories of pitch contour) beyond those involved in perceiving familiar talkers (those included in training) because perception of novel talkers cannot be accomplished using talker-specific strategies developed during training and, instead, must draw upon broad linguistic representations formed over talkers.

Method

Participants

A total of 160 participants (57 males; $M_{\text{age}} = 21.7$ years) provided data. All were paid for participation and gave informed consent. They were native speakers of English between the ages of 18 and 30 with normal or corrected-to-normal vision, no previous tone language experience, and no known hearing deficits.

Table 1 Distribution of amount of private music lessons reported by participants (in years, summed across all musical instruments)

Private lessons (years)	*n*	Percentage
None	62	38.8%
0–4	44	27.5%
5–9	21	13.1%
10–14	18	11.3%
15–19	7	4.4%
20–24	5	3.1%
25–29	2	1.3%
30–34	1	0.6%

All were recruited from the University of Maryland community. Although 226 participants began the study, 66 were excluded. Four were dismissed because of ineligible age (1), previous participation (1), or previous knowledge of a Chinese language (2). Nineteen others were excluded from the analyses because of heritage knowledge of a non-Chinese tone language. Additionally, 35 participants attrited before completing the study, and 8 were discarded due to missing data, failure to follow instructions, completion of training sessions at the wrong time intervals (i.e., more than 2 days apart or more than one session per day), and/or a lack of effort (e.g., impossibly quick response times suggesting that they were not attending to stimuli). Because musical experience and musical aptitude were variables of interest in this study, participants were recruited from the University of Maryland music department as well as the wider university community to ensure a broad range of musical experience. Of the final set of participants, 78% reported previous group music lessons, 61% reported previous private music lessons (see Table 1), and 12.5% were college music majors.

Materials

Participants completed five laboratory sessions, each lasting 1–2 hours.[5] No two sessions could be completed in the same day, and the fourth and fifth sessions had to occur within 2 days of one another. All sessions were held in a computer lab containing 14 individual computer stations separated by dividers. At any given session, 1–12 individuals participated, and proctors monitored participants' progress.

The task battery comprised a number of cognitive tests as well as a multisession training paradigm in which participants learned a set of Mandarin

Table 2 Predictor tasks by type and construct

Type	Construct	Task
Pitch ability	Tonal perceptual acuity	Tone discrimination
	Tonal perceptual acuity	Tone identification
	Pitch height STM	Pitch STM
	Pitch contour sensitivity	Pitch contour identification
Musicality	Musical aptitude	AMMA & parts of WMAT
	Musical experience	Musical background questionnaire
General L2 aptitude	Segmental perceptual acuity	Consonant discrimination
	Phonological STM	Nonword span
	WM – Updating	Running memory span
	WM – Inhibition	Antisaccade analogue
	L2 experience	Language background questionnaire
	Implicit induction	Serial reaction time
	Verbal associative memory	Paired associates
General cognitive ability	Explicit induction	Letter sets
	Crystallized intelligence	Wonderlic

pseudowords. The cognitive tests produced four types of predictor measures (see Table 2): (a) pitch ability measures hypothesized to be components of tonal aptitude, (b) musicality measures (incorporating both musical aptitude and musical experience), (c) control measures of general foreign language aptitude, and (d) control measures of general cognitive ability. Pitch ability measures (i.e., pitch-specific behavioral measures) were the primary focus of the study. The other measures were included to account for musicality, general cognitive ability, and overall L2 aptitude. Specifically, we wanted to determine how the pitch ability measures were related to other measures already known to have predictive validity for L2 learning and whether the pitch ability measures could provide additional predictive validity for learning of words that contrast crucially in tone.

Pitch Ability Tests

Four pitch-related tasks covered both linguistic and nonlinguistic tone processing as well as pitch height and pitch contour contrasts. Linguistic measures came from two perceptual tasks (discrimination and identification) involving isolated Mandarin tones.[6] Nonlinguistic measures were obtained using two perceptual tasks involving nonspeech pure tones (measuring STM for pitch height and identification of pitch contour).

Tone Discrimination

This task tested participants' ability to detect tonal differences using a categorial AX procedure (Flege, 2003). In AX discrimination, two auditory stimuli (A and X) are presented successively and the listener identifies them as either the "same" or "different." Stimuli in each trial of the current task contained the same segments, but sometimes different tones, and were produced by different talkers to prevent discrimination at a purely auditory level. The stimulus set consisted of 24 monosyllabic tonal minimal pairs, which involved 96 trials: 48 different trials (two presentations of each minimal pair in both orders) and 48 same trials. With an interstimulus and intertrial interval (ITI) of 1 second each, the task lasted 11–12 minutes and produced two measures: d' (a measure of perceptual sensitivity) and mean reaction time (RT) for accurate responses.

Tone Identification

An identification task was used in addition to a discrimination task because identification is more difficult (forcing listeners to make judgments about stimuli one at a time rather than in direct comparison) and more closely resembles ordinary speech processing. This task employed a one-interval, four-alternative forced-choice (4AFC) procedure in which the response options reflected the four-tone system of Mandarin (for examples of studies using this paradigm, see Wang, Spence, Jongman, & Sereno, 1999, and Wang & Kuhl, 2003). A 4AFC task is challenging due to multiple response options. To mitigate the difficulty, we eschewed arbitrary category labels in favor of graphic representations (line drawings) that transparently reflected the pitch contours of the tones. The task began with a 2-minute familiarization phase in which participants were exposed to the Mandarin tones and how they mapped onto the category labels. In the test phase, each trial presented a stimulus in isolation, which participants had to identify as one of the categories. Stimuli consisted of 20 monosyllabic tonal minimal quadruplets,[7] for a total of 80 test trials. With an ITI of 1 second, this task (including familiarization) lasted 9–10 minutes and produced two measures: percent accuracy and mean RT for accurate responses.

Pitch STM

This test measured auditory STM for nonspeech level tones (sine waves) in two conditions (e.g., Deutsch, 1970, 1972; Semal & Demany, 1993). In each condition, participants completed three practice trials with feedback before moving on to 48 test trials without feedback. In the first (control) condition, each trial involved participants hearing an alert sound indicating the trial's start, listening to a first tone, retaining it in memory during a 5-second silence, hearing a second tone, and then pressing a button to indicate whether the second tone was the same as or different from the first. In the following interference condition, trials were the same in structure except that during the 5-second interval between the two tones, six intervening tones were played. The design of this task, including the tone frequencies used, was modeled on Pechmann and Mohr (1992). This test took 10 minutes and produced two measures: percent accuracy in each of the two conditions.

Pitch Contour Identification

This test measured participants' sensitivity to pitch contour by having them identify the type of contour in 42 nonspeech tones (sine waves) varying in terms of pitch height and slope. Participants first completed 12 practice trials with feedback and then 336 test trials (four blocks of 84 trials) without feedback. On each trial, participants had to identify a stimulus as either "flat," "rising," or "falling" in pitch (for further details, see Bent, Bradlow, & Wright, 2006). This test took 30 minutes and produced two measures: percent accuracy and mean RT for correct responses.

Musicality Tests

Measures of musicality were obtained from two tests of musical aptitude and a questionnaire about previous musical experience.

Musical Aptitude

The primary test of musical aptitude was The Advanced Measures of Music Audiation (AMMA; Gordon, 1989), which requires participants to listen to pairs of short musical phrases and indicate whether the phrases are identical, different in pitch, or different in rhythm. This test took 20 minutes and was presented and scored by the AMMA software, yielding separate scores for Pitch and Rhythm. In addition, participants completed two parts of the Wing Musical Aptitude Test (WMAT; Wing, 1968), which provided information about perception of pitch in the context of a complex sound structure (a chord; WMAT, Part I) and memory for pitch in the context of a melody (WMAT, Part III). The WMAT took 10 minutes and produced an accuracy score for each part.

Musical Experience

Prior musical training and involvement were assessed through a musical background questionnaire covering both formal musical training and informal musical experience. The questionnaire included items from the Ollen Musical Sophistication Index (OMSI; Ollen, 2006) and Cuddy, Balkwill, Peretz, and Holden (2005), as well as items commonly collected in the literature, such as instruments played and length of private music lessons.

Foreign Language Aptitude Tests

Six tasks measuring aspects of L2 aptitude were included in the test battery. These tasks were motivated by recent research on L2 aptitude (Bowles, Silbert, Jackson, & Doughty, 2011; Bunting et al., 2011; Linck et al., 2012, 2013) and were used over alternatives such as the MLAT to examine pitch ability measures in the context of modern conceptualizations of L2 aptitude. The tasks were (a) Consonant Discrimination, which tested the ability to distinguish between two Hindi syllables evincing nonnative voicing contrasts; (b) Nonword Span, which measured phonological STM; (c) Running Memory Span, which measured WM; (d) Antisaccade Analogue, which measured the executive function of inhibition; (e) Serial Reaction Time, which measured implicit induction; and (f) Paired Associates, which measured verbal rote learning (patterned after a similar task in the MLAT). For more information about the structure and scoring of these tests, as well as their underlying cognitive constructs, see Doughty et al. (2010).

In addition to these cognitive tests, participants completed a language background questionnaire about early L2 experience and prior L2 study. Four measures were derived from this questionnaire: (a) whether or not the participant reported any heritage language exposure in the home before age 18 (Heritage Language Exposure), (b) the number of L2s previously studied (Number of L2s), (c) an average of the participant's self-reported highest reading and writing levels in any L2 (L2 Reading/Writing), and (d) an average of the participant's self-reported highest listening and speaking levels in any L2 (L2 Listening/Speaking). The latter two variables were measured on a 5-point rating scale (1 = *limited*, 5 = *excellent*).

General Cognitive Ability Tests

Each participant completed two control tasks measuring general cognitive ability (i.e., crystallized and fluid intelligence).

Wonderlic

The Wonderlic Contemporary Cognitive Ability Test, a revised version of the Wonderlic Personnel Test (Wonderlic Inc., 1999), measures general cognitive ability (g) and is highly correlated with longer tests of g such as the Wechsler

Adult Intelligence Scale. The test items are multiple-choice or short-answer questions of various types including vocabulary, arithmetic, and logical reasoning (Gesinger, 2001). The test was administered in a group setting, with a researcher controlling timing. The score was the total number of items successfully completed within 12 minutes.

Letter Sets

The Letter Sets Test is a short test of explicit induction drawn from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Derman, 1976) and was adapted for computer presentation in this study. The score was the total number of items successfully completed within 7 minutes.

Tonal Word Learning

To measure participants' ability to acquire the tone system of Mandarin, we used a word learning task (Tonal Word Learning) similar to that used in previous work on *ab initio* learning of Mandarin tone contrasts (Chandrasekaran et al., 2010; Wong & Perrachione, 2007; Wong et al., 2011). Over six sessions, participants learned to associate 24 tonal word forms (two monosyllabic tonal minimal quadruplets, four disyllabic tonal minimal quadruplets with tonal contrast on either the initial or final syllable) with unique meanings. The target items were pseudowords in the sense that none of the sound–meaning pairs represent real sound–meaning correspondences in Mandarin; the monosyllabic sound sequences occur, but have different meanings, whereas the disyllabic sound sequences do not occur (due to the given tone combination). This task provided the main dependent variable representing success at learning lexical tone: accuracy of meaning identification.

Although the basic design of our Tonal Word Learning task was similar to that used in several previous studies, it was most similar to that in Chandrasekaran et al. (2010) and Wong et al. (2011). In particular, our target lexicon comprised the same number of items (24) as in those two studies. Nevertheless, our target lexicon was more challenging to learn in two respects. First, it was more diverse phonologically, including not only monosyllabic items, but also longer (disyllabic) items; this required participants to learn more phonological material overall as well as to learn the tonal contrasts in a greater variety of contexts (in isolation, preceding another tone within the same word, and following another tone within the same word). Second, the nontonal (i.e., segmental) aspects of the target items differed to a greater degree from the segmental phonology of the L1 (English) because they were natural Mandarin syllables; that is, the items were not English-specific segmental sequences overlaid with Mandarin tone contours (cf. Wong & Perrachione, 2007), but Mandarin segmental sequences naturally produced with Mandarin tones.

In addition, the current Tonal Word Learning paradigm differed from that in Chandrasekaran et al. (2010) and Wong et al. (2011) in including fewer training sessions (six compared to more than nine). Because our objective was to provide enough training to be able to see individual differences in learning, we included only enough training sessions to be able to accomplish this while allowing the very best learners to reach ceiling. The number of sessions that accomplished this in Chandrasekaran et al. (2010) was 4–5; therefore, because our materials were more difficult, we included six sessions.[8] For logistical reasons, these six sessions were condensed into three visits to the laboratory (two sessions per visit).

The audiovisual stimuli for Tonal Word Learning comprised audio recordings of six native Mandarin speakers uttering the target phonological forms, as well as pictures representing the intended meanings. The six Mandarin speakers (three females; $M_{\text{age}} = 23.2$ years) were born and raised until at least the age of 18 in northern China with Mandarin as the primary language spoken at home; they were also proficient in English and were recorded in a sound-attenuated booth in the United States (using a head-mounted condenser microphone and mobile audio recorder, at 44.1 kHz and 24 bps). During recording, each of the target phonological forms was presented on an index card in terms of simplified Chinese characters and pinyin Romanization, and talkers were instructed to read the pinyin (due to the phonological ambiguity of many characters) at a comfortable volume and pace. To encourage natural production of the disyllabic items (which were all nonce forms by design), talkers were specifically instructed to say them normally, as if they were real words (i.e., without pausing between syllables). The intended meanings of the items were chosen using the MRC Psycholinguistic Database to be high in imageability, concreteness, and English frequency and were represented in the word learning task by colored line drawings (Rossion & Pourtois, 2004). For a full list of items, see Appendix S1 in the Supporting Information online.

Each session of Tonal Word Learning consisted of three interleaved phases: familiarization, training with feedback, and testing without feedback. In all phases, single-stimulus exposure to the stimuli was used because it has been shown to yield similar training benefits as paired-stimulus exposure (Wayland & Li, 2008) and corresponds closely with the listener's task in normal speech perception. Familiarization and training phases were blocked by tonal minimal quadruplet as in Wong et al. (2011), with monosyllabic quadruplets presented first. In all cases, the ITI was 1 second.

In each familiarization phase, participants were first exposed to the audiovisual stimuli (image depicting the target meaning on screen, audio recording of the word form over headphones) to be presented during the subsequent training phase. Each stimulus was presented four times, each time with the audio portion pronounced by four different talkers (two per sex). All familiarization phases combined lasted approximately 7–8 minutes. In each training phase, participants completed 16 trials of meaning identification with feedback. On each trial, they heard one of the four tonal word forms presented during the immediately preceding familiarization phase and tried to identify its meaning from among four response options (pictures representing the relevant minimal quadruplet). After entering their response, they were shown on screen whether they were correct or not. All training phases combined lasted 10–12 minutes. In each test phase, participants completed 96 trials of meaning identification (four presentations of each of the 24 stimuli) without feedback. Test trials were identical to training trials except that they provided no feedback, were mixed (rather than blocked by quadruplet), and presented all 24 response options. The test phase lasted 9–10 minutes. In the first five test phases, the audio stimuli were the recordings presented during training, whereas in the sixth test phase, they were new recordings uttered by unfamiliar talkers.

Tonal Word Learning produced accuracy measures for each of the six test phases, and the final two of these were analyzed: Penultimate Accuracy (accuracy during the fifth test phase with trained stimuli) and Final Accuracy (accuracy during the sixth test phase with untrained stimuli from novel talkers). Final Accuracy was analyzed to examine the ability to transfer learning from familiar to unfamiliar talkers. Absolute chance performance on both measures was equivalent to 4.2% (1/24). If the participant could narrow down the answer choices to the four in the correct minimal quadruplet, however, then chance performance increased to 25%.

Procedure

During the first day of the study, participants gave informed consent, listened to instructions, and completed the Wonderlic. Next, they were familiarized with the computer and testing equipment. For the latter half of this day (and in all subsequent days), participants worked at their own pace through a written list of assigned tasks (including optional breaks). All participants completed the tasks in a fixed order. The first 2 days consisted of predictor tasks from the test battery. The final 3 days each required participants to complete two sessions of Tonal Word Learning, which were separated by one or more intervening tasks measuring relatively stable (long-term) individual differences not directly related to Mandarin and thus unlikely to be affected by a few sessions of Tonal Word Learning. These tasks were the WMAT, Consonant Discrimination, Antisaccade Analogue, Serial Reaction Time, and Letter Sets. Most tasks were presented via personal computer running E-Prime 2.0. Exceptions were the AMMA (presented on computer via CD), the Wonderlic (given in paper-and-pencil form), and the two questionnaires (paper-and-pencil). Some computerized tasks required participants to use a response box; others required a mouse or keyboard. During testing, participants wore headphones.

Results

Data Preparation

For RT variables, data were examined by participant for each task and any data points greater than three standard deviations from that participant's mean for that condition were replaced with the value at three standard deviations from the mean. Then, overall distributions were examined for each variable, and transformations (arcsine transformation for percent accuracy, log transformation for RT; e.g., Ratcliff, 1993) were applied if they improved normality. To reduce the number of variables considered during modeling, Musical Background Questionnaire items were subjected to a principal component analysis (PCA). The first component that emerged in this analysis captured 34% of the variance in questionnaire responses and was judged on the basis of the loadings to best represent overall musical experience. The three musical experience variables that loaded most strongly on this first component were selected to be carried forward: (a) overall score on the OMSI (OMSI Score),[9] (b) whether or not the participant was a college music major (Music Major), and (c) total months of private music lessons (Months of Private Music Lessons), which was calculated by summing over all instruments (even when lessons for multiple instruments occurred during overlapping periods of time).

PCA of the Predictors

To examine the relationships among predictors and to further inform data reduction decisions, measures of pitch ability, musicality (including the three selected musical experience variables), general L2 aptitude (including the four variables derived from the language background questionnaire), and general cognitive ability were entered into a PCA. Principal axis factoring was used to extract components with eigenvalues greater than 1, and for ease of interpretation, the solution was rotated using the direct oblimin rotation, a type of oblique rotation (Harman, 1976). This procedure resulted in an eight-component solution (for loadings, see Appendix S2 in the Supporting Information online).

Component I: Musical Experience

This component accounted for 24.2% of variance in the correlation matrix of the predictor variables.[10] All three musical experience variables loaded strongly on this component.[11] In addition, accuracy on both parts of the WMAT and Pitch STM–Interference loaded moderately. Thus, this component may represent prior musical training and memory for nonlinguistic tones in the context of competing tones.

Component II: RT

This component accounted for 9.2% of variance. RT measures from four tasks (Tone Identification, Tone Discrimination, Pitch Contour Identification, Antisaccade Analogue) loaded strongly or moderately on this component, separately from their respective accuracy measures. Thus, this component represents RT and indicates that the RT and accuracy measures from these tasks represent distinct dimensions.

Component III: Foreign Language Experience

This component accounted for 7.9% of variance. Two variables related to L2 attainment (L2 Reading/Writing, L2 Listening/Speaking) loaded strongly and Heritage Language Exposure loaded moderately on this component, suggesting that it represents prior L2 experience.

Component IV: Verbal Memory/Implicit Induction

This component accounted for 6.8% of variance. Measures of verbal associative memory and phonological STM (from Paired Associates and Nonword Span) loaded strongly, while a measure of implicit induction (from Serial Reaction Time) loaded moderately. These results suggest that this component represents the verbal memory and implicit induction aspects of L2 aptitude.

Component V: Musical Aptitude

This component accounted for 5.4% of variance. Accuracy on both sections of the AMMA loaded strongly on this component, suggesting that this component captures musical aptitude as operationalized by that specific test.[12]

Component VI: Pitch Processing

This component accounted for 4.8% of variance among the predictor variables. Accuracy on Pitch STM–Control loaded strongly on this component, and accuracy in several additional pitch ability tasks loaded moderately: Tone Discrimination, Tone Identification, Pitch Contour Identification, and the WMAT, Part III (testing pitch memory). This component thus seems to encompass aspects of both linguistic and nonlinguistic pitch processing, providing further evidence of joint processing (e.g., Perrachione et al., 2013).

Component VII: Inhibition

This component accounted for 4.2% of variance. The accuracy and RT measures from Antisaccade Analogue (testing inhibition) loaded together on this component. Consequently, this component appears to be related to inhibitory control.

Component VIII: General Learning Ability

The final component accounted for 4.0% of variance. No measures loaded strongly on this component. Loading moderately were measures reflecting general cognitive ability (from Letter Sets and Wonderlic) and pitch ability (accuracy on Tone Identification, which involved learning to label pitch patterns). Together, these loadings suggest that this component represents general learning or reasoning ability.

Regression Analyses of Learning

Whereas the PCA results provided information about the relationship of measures within the set of potential predictors, regression analyses were needed to determine which measures were most predictive of success in Tonal Word Learning. Mean accuracies in the penultimate and final test phases of Tonal Word Learning were similar (penultimate: $M = 53\%$, $SD = 25\%$; final: $M = 52\%$, $SD = 25\%$; see Appendix S3 in the Supporting Information online for further descriptive statistics) and showed significant learning. Although the mean accuracy in the final test phase (Session 6) was only 52%, this represented a 48% increase over chance performance (4%), indicating that substantial learning had occurred by the end of training. Learning was also apparent in the overall trajectory of accuracy during training (Figure 1), which steadily increased from Session 1 to Session 5 before decreasing slightly in Session 6 (the generalization session introducing novel talkers at test); this was also the case when the data were split into "good learners" above the median in Session 6 ($n = 78$) and "poor learners" below the median in Session 6 ($n = 78$), although good learners showed a steeper trajectory.

The likelihood of an accurate response in the penultimate and final test phases (Penultimate Accuracy and Final Accuracy, respectively) was analyzed in separate mixed-effects logistic regression models (e.g., Jaeger, 2008), starting with random effects for Participant and Quadruplet (i.e., the tonal minimal

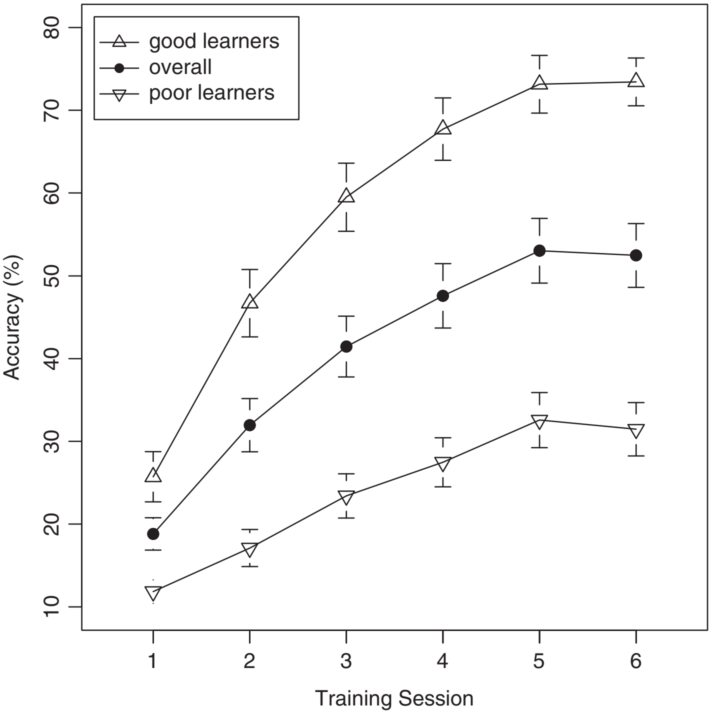

Figure 1 Percent accuracy at test over the six sessions of Tonal Word Learning. Means are plotted over all 160 participants, over participants above the median in Session 6 (good learners), and over participants below the median in Session 6 (poor learners). Error bars represent 95% confidence intervals of the mean over participants.

quadruplet represented by an item) and adding in fixed effects for the predictor variables. In all, there were 22 possible fixed-effect predictors (6 for pitch ability, 7 for musicality, 7 for general L2 aptitude, 2 for general cognitive ability); see Appendix S4 and Appendix S5 in the Supporting Information online for the correlation matrix of these predictors and reliability statistics. Because our research questions all concerned pitch ability, all nine accuracy measures from pitch-related behavioral tasks and all three of the selected musical experience measures were retained for modeling. As discussed above, RT measures loaded on a component together, so only the strongest-loading measure (from Tone Discrimination) was carried forward as the final pitch ability measure. Because the L2 experience variables also loaded together, only the strongest-loading (L2 Listening/Speaking) was carried forward as a L2 aptitude measure, along with the six behavioral L2 aptitude measures. Finally, measures from both Letter Sets and Wonderlic were carried forward as general cognitive ability measures.

Considering the relatively small ratio of participants to predictors, we followed a two-step process in modeling. First, single-predictor models were run for each predictor to rank the predictors according to informativeness as indexed by the Akaike information criterion (AIC; Akaike, 1974). Second, full models were built incrementally in order to avoid overfitting the data, and at each stage a predictor was kept only if it significantly improved the model's predictions according to likelihood-ratio tests. Order of entry was determined by informativeness; that is, the predictor that was most informative on its own (associated with the lowest AIC in a single-predictor model) was entered first, followed by the next most informative predictor, and so on. This method of entry was used to provide a strong test of whether a given predictor should be kept in the model.

Six predictors significantly improved the model of Penultimate Accuracy beyond random effects: by order of entry, accuracy in Tone Identification, *d′* in Tone Discrimination, Nonword Span, Paired Associates, Months of Private Music Lessons, and Letter Sets, all $\chi^2(1) > 8.29, p < .001$. No other variables further improved the model, all $\chi^2(1) < 3$. The fixed-effect coefficients for this model are given in Table 3. All are positive, indicating that higher scores on each of the predictor measures were associated with a higher likelihood of accuracy.

Eight predictors significantly improved the model of Final Accuracy beyond random effects: by order of entry, accuracy in Tone Identification, *d′* in Tone Discrimination, Nonword Span, Pitch Contour Identification, Letter Sets, Pitch STM—Interference, Paired Associates, and Months of Private Music Lessons, all $\chi^2(1) > 5.01, p < .05$. Most of these were also predictive of Penultimate Accuracy; however, Pitch Contour Identification and Pitch STM—Interference were not. Thus, two additional predictors beyond those in the model of Penultimate Accuracy emerged as significant in this model. No other measures further improved the model, all $\chi^2(1) < 3$. The fixed-effect coefficients are shown in Table 3; all are again positive, indicating that higher scores on the predictor measures were associated with a higher likelihood of accuracy.

Individual Differences in Learning

To explore individual differences in learning in greater detail, the learning trajectory of each participant was plotted according to quartile in the final test phase (Figure 2). Examination of these individual trajectories revealed that, as in Chandrasekaran et al. (2010) and Wong et al. (2011), there was considerable

Table 3 Coefficients of significant fixed-effect predictors in final models for penultimate accuracy and final accuracy

Outcome	Predictor	*β* (*SE*)	*z*
Penultimate accuracy	(Intercept)	–1.82 (0.34)	–5.29***
	Tone identification (accuracy)	2.15 (0.31)	7.01***
	Tone discrimination (*d′*)	0.26 (0.12)	2.17*
	Nonword span	0.14 (0.08)	1.81†
	Paired associates	0.44 (0.12)	3.83***
	Months of private music lessons	0.01 (0.01)	3.67***
	Letter sets	0.26 (0.09)	2.92**
Final accuracy	(Intercept)	–3.23 (0.58)	–5.60***
	Tone identification (accuracy)	1.81 (0.29)	6.18***
	Tone discrimination (*d′*)	0.30 (0.12)	2.64**
	Nonword span	0.19 (0.07)	2.72**
	Pitch contour identification	0.95 (0.56)	1.70†
	Letter sets	0.26 (0.08)	3.16**
	Pitch STM – Interference	0.02 (0.01)	2.75**
	Paired associates	0.33 (0.11)	3.10**
	Months of private music lessons	0.01 (0.01)	2.26*

Note. The listed order of predictors reflects order of entry (determined on the basis of Akaike information criterion in a single-predictor model). $^{\dagger}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

variation among participants with respect to their starting and ending points, rate of learning, and steadiness of improvement. Consistent with the differences between good learners and poor learners seen in Figure 1, participants varied widely in terms of their ultimate accuracy in Session 6; however, they also varied in terms of their initial accuracy in Session 1, as shown in Figure 2.

In addition, there were marked differences among participants in their rate of learning (i.e., the slope of their trajectory). In some cases, the slope of the trajectory could be attributed to the participant's starting point, but in general the differences in slope occurred in spite of differences in starting point. Participants with the lowest Session 1 accuracies (mostly in the first and second quartiles) generally showed the flattest trajectories despite having the most room for improvement, whereas participants in the fourth quartile showed the steepest trajectories (especially between Session 1 and Session 3, the midpoint of the learning regimen). In this regard, there was a notable contrast between participants in the first and second quartiles: Although these two groups showed similar levels of accuracy in Session 1, the second-quartile

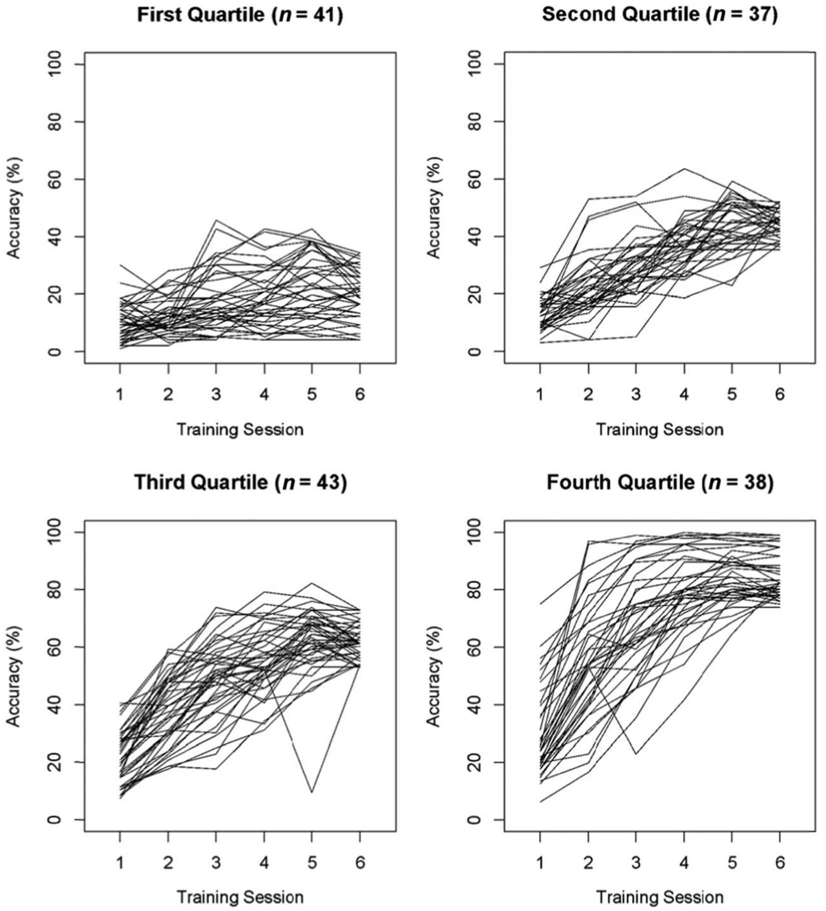

Figure 2 Individual trajectories in percent accuracy at test over the six sessions of Tonal Word Learning, according to quartile in Session 6 (one participant is excluded from the third quartile due to missing data from Session 1).

group showed considerably more learning over time. Compared to learning trajectories in the second quartile, those in the third quartile tended to be steeper, and those in the fourth quartile, steeper still. Participants in the fourth-quartile group tended to show particularly fast rates of learning, such that some were already performing at ceiling in Sessions 2 and 3. Differences between quartiles in rate of learning were reflected in significantly greater gains in accuracy from Session 1 to 3 for each of quartiles 2–4 in comparison to the quartile immediately below, all Welch-corrected two-sample ts > 3.66, $p < .001$.

Finally, participants also showed variation in the stability of their learning gains. Although the general pattern across sessions was for accuracy to increase, many participants showed dips in accuracy between consecutive sessions, which in some cases were quite large. On the whole, however, decreases in accuracy between consecutive sessions were relatively small compared to increases in accuracy between consecutive sessions, suggesting that most participants were indeed learning from the training, even if their progress over time was susceptible to temporary setbacks.

Discussion

In this study, we explored three main questions. First, can measures of pitch ability and musicality predict learning of L2 tonal contrasts once we control for general cognitive ability and general L2 aptitude? That is, can we identify an aptitude for the tonal aspect of a L2 specifically? Second, which aspects of pitch ability are central to such an aptitude for tone? Third, how do measures of pitch ability relate to one another and to measures of musicality, general cognitive ability, and general L2 aptitude?

Results were consistent with all three of our hypotheses. First, measures of linguistic pitch processing and musicality improved predictions of Tonal Word Learning performance beyond general cognitive ability and L2 aptitude measures, suggesting that aptitudes for specific L2 features such as tone can indeed be identified. Second, although one measure of musical experience was a significant predictor of Tonal Word Learning performance, the strongest predictors were behavioral tests of pitch processing, especially linguistic tasks such as Tone Discrimination and Tone Identification, similar to Cooper and Wang's (2012) findings for Cantonese. Because these two tasks employed the Mandarin tone system, it is not surprising that they were the best predictors of success in a Mandarin learning task; in this sense, such language-specific tasks may be similar to "work sample tests" such as those explored by Carroll and Sapon (1955) during development of the MLAT. Finally, transfer of learning to new talkers was related to additional predictors beyond those that predicted performance on talkers included in the training set (namely, Pitch Contour Identification and Pitch STM). Consistent with previous studies (e.g., Wong & Perrachione, 2007), this suggests that nonverbal aspects of pitch processing are important for generalization of tone learning by L1 speakers of nontonal languages. That is, aspects of both STM for pitch and judgment of linguistic pitch contour may be particularly important for the process of transferring categorical information about the learned Mandarin tone system onto judgments of new talkers, who differ in pitch register, pitch range, and other vocal characteristics. Interestingly, musical experience predicted success only with familiar talkers and not with new talkers, suggesting that, although musicality may be important for initial learning of lexical tone contrasts, it may be less relevant for transfer of learning to new talkers once an initial level of competence with the training set has been established.

Although we had no strong a priori hypothesis regarding how the pitch ability measures would relate to one another or to other measures from the battery, our analyses suggested that some behavioral measures of pitch processing ability were separable from musical experience measures. In doing so, the current analyses move beyond previous work on the relationship of music and lexical tone learning (e.g., Wong & Perrachione, 2007) by providing new information about the shared variance among musicality measures and other measures of pitch ability. Our results also indicated that both the pitch processing and musical experience components were distinct from general cognitive ability and from aspects of L2 aptitude. Measures that loaded most strongly on the pitch processing component were, in turn, the most informative predictors of Tonal Word Learning performance. Additionally, the results suggested that different tests of musical aptitude may be measuring quite different aspects of musical processing or experience.

Given the concern of previous studies with the relationship between music and lexical tone, we included several music-related predictor variables in this study; notably, however, with only one exception (the experiential variable Months of Private Music Lessons) these failed to provide predictive validity beyond that of tone perception measures. This is notable because in contrast to previous studies, which were limited to group-level comparisons of musicians versus nonmusicians or good versus poor pitch perceivers (e.g., Cooper & Wang, 2012; Wong & Perrachione, 2007), the large number of participants in the current study enabled a more sensitive correlational treatment of musicianship in terms of a continuum of individual differences. With this treatment, musical variables on the whole did not turn out to be powerful predictors of tonal word learning compared to tone-specific measures. These results suggest that previous findings of an advantage for musicians in tone-related tasks are driven in large part by underlying individual differences in overall pitch sensitivity and processing that are correlated with previous musical experience (e.g., Perrachione et al., 2013). This correlation could arise because musical training enhances pitch ability and/or because people with high levels of pitch ability tend to be the ones who gravitate toward musical training and thus tend to be selected into studies of musicians. Whatever the source of this correlation, it is clear that musical variables, which are only indirectly related to tone, are less effective predictors of tonal word learning than measures directly related to tone perception.

Although it may seem obvious that tone perception measures should be the best predictors of tonal word learning, this finding serves to support our larger point about the relationship between L2 aptitude and L2 attainment—namely, that a feature-specific approach to the prediction of L2 attainment (drawing on abilities/aptitudes that are most closely related to the given linguistic challenge) is more powerful than a language-general approach (e.g., based on general L2 aptitude tests such as those summarized in the introduction section). This point converges with the observation of Perrachione et al. (2011) that a more domain-specific behavioral measure predicted tonal word learning better than a more general measure did. To reiterate, we argue that acquisition outcomes can be predicted more effectively through careful consideration of the specific challenges presented by a given L2 or type of L2, along with the specific skills required to meet those challenges. Thus, the contribution of the current study is in demonstrating that mastery of a feature of a target language known to be particularly challenging for L2 learners—as a necessary component of learning the language at large—is predicted most successfully by behavioral measures that are most relevant to that feature. The validity of this feature-specific approach to predicting L2 attainment is important to establish because it remains to be widely acknowledged in the literature.

Although the present study takes a promising first step toward identifying aptitudes for specific L2 features, there are some limitations that should be acknowledged. First, the generalizability of the results is constrained by the laboratory-based nature of the study and by the restricted nature of the outcome measures. For example, it is not clear whether the results reported here would be found in more naturalistic L2 learning situations, with more top-down contextual information, or when using production measures. In addition, the confines of our word learning task did not allow for a systematic examination of other features of tone systems (e.g., tone sandhi, secondary phonetic cues such as duration and voice quality) or, for that matter, other tone systems besides that of Mandarin (in particular, register tone systems where the tonal contrasts center around pitch height rather than pitch contour). Finally, only cognitive and perceptual measures were included in the test battery; consequently, the influence of other potentially relevant predictors of tonal word learning (such as motivational factors and personality characteristics) is unknown.

Future Research and Conclusions

This work provides a proof of concept for the study of aptitudes underlying success with specific L2 features. That is, the current findings demonstrate that skills relevant to a challenging aspect of a L2 may provide important sources of information regarding individual differences in learning of that L2. As such, this area represents fertile ground for future research on L2 aptitude. Whereas differential prediction of L2 aptitude has been difficult to establish for individual L2s or classes of L2s (due to theoretical and practical constraints), focusing on specific typological features as we have done in this study may allow for greater progress in identifying specific, rather than general, aptitudes for L2 learning. Incorporating these types of feature-specific aptitude measures into existing test batteries may lead to improved prediction of learner success, with a number of practical consequences. For example, such improved prediction could have important policy implications for government training programs or similar situations involving selection of learners into languages. More generally, greater precision in the evaluation of aptitude as it relates to a given L2 would provide both teachers and students with more useful information on the challenges of tackling the given L2 for particular students, enabling early intervention through additional training in specific skills.

Several avenues for future research are suggested by the current findings. The next logical step would be to test whether the predictors of tonal aptitude identified here, which are hypothesized to be general predictors of tone learning, do in fact predict tone learning generally. Insofar as pitch ability underlies tone learning in all tone languages, we might expect these findings to generalize to other tone languages. To what extent the findings do generalize—in particular, to tone languages with qualitatively different tone systems (e.g., multiple level tones)—is an empirical question that remains to be tested. Additional work should also examine whether the predictors of tone learning success found here transfer to naturalistic learning conditions and to later stages of learning.

Although the general approach to aptitude we have taken in this study has been applied here to a feature of L2 phonology, it is not difficult to see how it could be extended to other kinds of L2 features. For example, whereas existing L2 aptitude tests focus on general learning ability or on auditory perception and processing, the learning of new orthographies (visual representations of language) might rely heavily on visuospatial processing skills, which are not generally examined as part of L2 aptitude. Consequently, for a L2 with a complex orthography, might it be the case that the ease of acquiring literacy (i.e., orthographic aptitude) is more effectively predicted by a learner's visuospatial processing skills than by general L2 aptitude measures? This is the type of question that follows naturally from a featural approach to L2 aptitude. Thus, this approach to aptitude research provides a framework for formulating concrete research questions and, ultimately, developing more fine-grained knowledge of the relationship between individual differences and L2 learning success.

Final revised version accepted 24 July 2015

Notes

1 "VORD" is not an acronym. It is the word for "word" in the artificial language used in the test.

2 Generally, musical experience refers to past musical training, whereas musical aptitude refers to the potential for success with future musical training (which is typically measured via an aptitude test). However, the relationship between these two constructs has not been well defined in the literature.

3 Brutten et al. (1985), Dexter (1934), Fenner (1955), Leutenegger and Mueller (1964), and Pimsleur et al. (1962) used the Seashore Measures of Musical Talents (Seashore, 1960); Slevc and Miyake (2006) used the Wing Musical Aptitude Test (Wing, 1968); and Gilleece (2006) used the Bentley Musical Aptitude Test (Bentley, 1966).

4 By general cognitive ability, we mean both crystallized and fluid intelligence. Because our dependent measure is a learning task, including such measures provides a strong test of whether targeted measures of tonal aptitude are capturing variance in tone learning beyond that explained by individual differences in general learning speed or efficiency.

5 Forty-nine participants also completed 1–2 additional sessions comprising a speech production task, electroencephalography (EEG), and/or magnetic resonance imaging (MRI) scans. Participants were selected into these additional sessions if they volunteered and met additional criteria. Results from these sessions are not reported here. The MRI sessions did not involve any stimuli. The EEG sessions, however, involved listening to tones and thus could have affected tonal word learning. To examine this possibility, an independent-samples t test was conducted comparing the EEG participants ($n = 32$) to the non-EEG participants ($n = 128$) on accuracy during the first training session. Although EEG participants performed slightly better (22%) than non-EEG participants (17%), this difference was not significant, $t(156) = -1.64$, $p = .10$, suggesting that the additional session did not significantly affect tonal word learning.

6 A third tone perception task, which measured the ability to categorize sentence-embedded tones from a different L2 (Burmese), was included in the test battery for a different purpose and is not reported here. However, the regression analyses reported below were also attempted with the measure from this task (percent accuracy) in the pool of predictors, and the measure did not emerge as a significant predictor.

7 A tonal minimal quadruplet is a set of four words that share the same segments, but differ in tone and, therefore, meaning (e.g., [ma] uttered with each of the four Mandarin tones).

8 It should be noted that, despite our significantly more difficult training paradigm, the performance of our 160 learners was overall quite close to that of the 16 learners in Chandrasekaran et al. (2010). Mean accuracy in Session 5 (pregeneralization) was around 53% in the current study versus around 58% in Chandrasekaran et al. (2010), judging from their Figure 2.

9 Although some subscores on the OMSI had overall higher component loadings than the total OMSI score, the sum score was carried forward as it was judged to most reliably reflect various aspects of musical experience.

10 The determination of which predictor variables account for the most variance in tonal word learning performance (the outcome variable) is addressed below.

11 In these summaries, “strongly” refers to loadings of .65 or higher, and “moderately” refers to loadings between .35 and .64. For the exact component loadings, see Appendix S2 in the Supporting Information online.

12 Although the WMAT has been used as a musical aptitude measure, it loaded moderately onto the musical experience component. This may be because performance on the WMAT subtasks used (indicate the number of notes in a chord; indicate the note changed in longer sequences) is more susceptible to influence from musical training (experience) than performance on the AMMA.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. doi:10.1017/S027226310808073X

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

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

Asher, J. J. (1972). Differential prediction of student success in intensive language learning. ERIC Document Reproduction Service No. ED066986.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 97–103. doi:10.1037/0096-1523.32.1.97

Bentley, A. (1966). *Measures of music abilities*. Berkshire, UK: NFER-Nelson Publishing Company.

Bidelman, G., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLOS One*, *8*, e60676. doi:10.1371/journal.pone.0060676

Bowles, A. R., Silbert, N. H., Jackson, S. R., & Doughty, C. J. (2011). *Individual differences in the ability to discriminate foreign language phonological contrasts*. Poster presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA.

Brutten, S. R., Angelis, P. J., & Perkins, K. (1985). Music and memory: Predictors for attained ESL oral proficiency. *Language Learning*, *35*, 299–313. doi:10.1111/j.1467-1770.1985.tb01030.x

Bunting, M. F., Bowles, A. R., Campbell, S. G., Linck, J. A., Mislevy, M. A., Jackson, S. R., et al. (2011). *Reinventing DLAB: Potential new predictors of success at DLIFLC*. Technical Report. University of Maryland Center for Advanced Study of Language.

Carroll, J. B. (1958). Factor analysis of two foreign language aptitude batteries. *Journal of General Psychology*, *59*, 3–19. doi:10.1080/00221309.1958.9710168

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Carroll, J. B., & Sapon, S. M. (1955). Prediction of success in a work-sample course in Mandarin Chinese. *American Psychologist*, *10*, 492–493.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test, Form A*. New York: The Psychological Corporation.

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

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

Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. *Annals of the New York Academy of Sciences*, *1060*, 311–324. doi:10.1196/annals.1360.026

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, *7*, 203–207. doi:10.1007/s10339-006-0146-7

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*, 46–61. doi:10.1080/09541440802708136

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*, 1604–1605. doi:10.1126/science.168.3939.1604

Deutsch, D. (1972). Effect of repetition of standard and of comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, *93*, 156–162. doi:10.1037/h0032496

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception: An Interdisciplinary Journal*, *21*, 339–356. doi:10.1525/mp.2004.21.3.339

Dexter, E. S. (1934). Pitch discrimination and French accent on the high school level. *Journal of Applied Psychology*, *18*, 717–720. doi:10.1037/h0072734

Dillon, S. (2010, January 20). Foreign languages fade in class—except Chinese. *The New York Times*. Retrieved from http://www.nytimes.com/2010/01/21/education/21chinese.html

Doughty, C., Campbell, S., Bunting, M., Mislevy, M., Bowles, A., & Koeth, J. (2010). Predicting near-native L2 ability: The factor structure and reliability of Hi-LAB. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected proceedings of the 2008 Second Language Research Forum: Exploring SLA perspectives, positions, and practices* (pp. 10–31). Somerville, MA: Cascadilla Press.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Fenner, R., Jr. (1955). *An exploratory study of some variables for measuring ability of students entering college with diversified training in French to learn to speak the language*. Unpublished doctoral dissertation. Syracuse University, New York.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–43). Padova, Italy: Unipress.

Gesinger, K. F. (2001). Review of the Wonderlic Personnel Test and Scholastic Level Exam. In B. S. Blake & J. C. Impara (Eds.), *The fourteenth mental measurements yearbook*. Retrieved from http://buros.org

Gilleece, L. F. (2006). *An empirical investigation of the association between musical aptitude and foreign language aptitude*. Unpublished doctoral dissertation, University of Dublin, Trinity College, Ireland.

Gordon, E. (1989). *Advanced measures of music audiation*. Chicago: GIA Publications.

Gottfried, T. L., & Ouyang, G. Y.-H. (2005). Production of Mandarin tone contrasts by musicians and non-musicians. *Journal of the Acoustical Society of America*, *118*, 2025. doi:10.1121/1.4785767

Gottfried, T. L., & Ouyang, G. Y.-H. (2006). Training musicians and nonmusicians to discriminate Mandarin tones. *Journal of the Acoustical Society of America*, *120*, 3167. doi:10.1121/1.4787905

Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America*, *115*, 2545. doi:10.1121/1.4783674

Harley, B., & Wang, W. (1997). The critical period hypothesis: Where are we now? In A. M. B. D. Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 19–51). Mahwah, NJ: Erlbaum.

Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99. doi:10.1016/0010-0285(89)90003-0

Lett, J. A., & O'Mara, F. E. (1990). Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defense Language Institute Foreign Language Center. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 222–260). Englewood Cliffs, NJ: Prentice Hall Regents.

Leutenegger, R. R., & Mueller, T. H. (1964). Auditory factors and the acquisition of French language mastery. *The Modern Language Journal*, *48*, 141–146. doi:10.1111/j.1540-4781.1964.tb04478.x

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., et al. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566. doi:10.1111/lang.12011

Linck, J. A., Jackson, S. R., Silbert, N. H., Tare, M., Bowles, A., Campbell, S. G., et al. (2012). *Individual differences in executive functions predict second language learning success*. Poster presented at the 53rd Annual Meeting of the Psychonomic Society, Minneapolis, MN.

MacWhinney, B. (1996). Language-specific prediction in foreign language learning. *Language Testing*, *12*, 292–320. doi:10.1177/026553229501200303

Maddieson, I. (2013). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/13

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

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*, 15894–15898. doi:10.1073/pnas.0701498104

Neumann, I., Abrahams, N. M., & Githens, W. H. (1968). *Selection of military personnel for foreign language training*. Naval Personnel Research Activity San Diego, CA Rep. No. Z998976. Retrieved from http://stinet.dtic.mil/oai/oai?&verb=getRecord&metadataPrefix=html&identifier=AD0679899

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*. Unpublished doctoral dissertation, Ohio State University, Ohio.

Parry, T. S., & Child, J. R. (1990). Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 30–66). Englewood Cliffs, NJ: Prentice Hall Regents.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 1–14. doi:10.3389/fpsyg.2011.00142

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, *20*, 314–320. doi:10.3758/BF03199668

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLoS ONE*, *8*, e73372. doi:10.1371/journal.pone.0073372

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

Petersen, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement*, *36*, 369–380. doi:10.1177/001316447603600216

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt Brace Jovanovich.

Pimsleur, P., Stockwell, R. P., & Comrey, A. L. (1962). Foreign language learning ability. *Journal of Educational Psychology*, *53*, 15–26. doi:10.1037/h0044336

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532. doi:10.1037/0033-2909.114.3.510

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*, 368–392. doi:10.1177/026765830101700405

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. doi:10.1017/S0267190505000036

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236. doi:10.1068/p5117

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349. doi:10.1111/1469-8986.00172.x

Seashore, C. (1960). *Seashore measures of musical talent*. New York: The Psychological Corporation.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, *94*, 1315–1322. doi:10.1121/1.408159

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam: John Benjamins.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*, 675–681. doi:10.1111/j.1467-9280.2006.01765.x

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of second language aptitude and second-language proficiency. *The Modern Language Journal*, *95*, 253–273. doi:10.1111/j.1540-4781.2011.01176.x

Starr, D. (2009). Chinese language education in Europe: The Confucius Institutes. *European Journal of Education*, *44*, 65–82. doi:10.1111/j.1465-3435.2008.01371.x

Tanaka, A., & Nakamura, K. (2004). Auditory memory and proficiency of second language speaking: A latent variable analysis approach. *Psychological Reports*, *95*, 723–734. doi:10.2466/pr0.95.3.723-734

Tierney, A. T., & Pisoni, D. B. (2003). Some effects of early musical experience on sequence memory spans. *Research on Spoken Language Processing*, *26*, 289–305.

Wang, Y., & Kuhl, P. K. (2003). Evaluating the "critical period" hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10, and 14 years of age. In M. J. Solé & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1537–1540). Barcelona, Spain: Futurgraphic.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658. doi:10.1121/1.428217

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*, 654–662. doi:10.1016/j.wocn.2010.10.001

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*, 250–267. doi:10.1016/j.wocn.2007.06.004

Wing, H. D. (1968). *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity* (2nd ed.). London: Cambridge University Press.

Winke, P. (2007). Tuning into tones: The effect of L1 background on L2 Chinese learners' tonal production. *Journal of the Chinese Language Teachers Association*, *42*, 21–55.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, *97*, 109–130. doi:10.1111/j.1540-4781.2013.01428.x

Wonderlic, Inc. (1999). *Wonderlic personnel test*. Vernon Hills, IL: Author.

Wong, F. C. K., 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*, 8780–8785. doi:10.1523/JNEUROSCI.0999-11.2011

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. doi:10.1017/S0142716407070312

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422. doi:10.1038/nn1872

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Target Lexicon in the Tonal Word Learning Task.
Appendix S2. PCA Component Loadings.
Appendix S3. Descriptive Statistics for Predictor and Dependent Measures.
Appendix S4. Correlation Matrix of Predictor Measures.
Appendix S5. Reliability of Predictor Measures.