

RESEARCH

Open Access



The role of novelty stimuli in second language acquisition: evidence from the optimized training by the Pinyin Tutor at TalkBank

Yanhui Zhang^{1*} and Brian MacWhinney²

*Correspondence:
Yanhui.Zhang@nottingham.
edu.cn

¹ University of Nottingham
Ningbo China, Ningbo, China

² Carnegie Mellon University,
Pittsburgh, USA

Abstract

As hypothesized by the unified competition model (MacWhinney, 2007, 2017, 2021), optimizing training schemes can enhance second language (L2) learning by fostering various protective factors. Under such a framework, the current study focuses on how the familiarity of stimuli will affect learning Chinese phonetic skills in a computer-assisted language learning (CALL) environment. Two training conditions, i.e., training with familiar stimuli from the textbook and unfamiliar stimuli from novelty design, were administered for two groups of learners at American universities, where the classroom instructions were integrated with the Pinyin Tutor—an online spoken Chinese learning platform hosted under TalkBank. The results show that training with novelty stimuli leads to a greater pretest–posttest improvement for intermediate learners, whereas more significant improvement has been observed in training with familiar stimuli among beginning learners. The learning-enhancing power of the Pinyin Tutor is evidenced by the overall significance of the pretest–posttest improvement when consolidating the results of the two conditions. Furthermore, high retention has been demonstrated in all six aspects of the Pinyin knowledge as tested by a three-month-after delayed posttest. These findings tend to endorse a differentiated design of instructional materials with increasing novelty components as the level of L2 learning advances. The overall significant learning-boosting results accredit the design of the Pinyin Tutor, where the technological architecture and algorithms were integrated with psycholinguistic and pedagogical theories. Suggestions and implications for smart learning in general are presented.

Keywords: Optimal training, Novelty stimuli, Cue familiarity, Phonetic skill, L2 Chinese, L2 speech perception, Pinyin Tutor, Unified competition model

Introduction

Research in the domain of second language acquisition has generally found that adults are inferior to infants or children in the ability to perceive and produce novel foreign speech sounds. It is challenging for adult second language learners to perceive the speech contrasts that do not exist in their native language. A frequently cited example

is for non-native speakers to learn Chinese Pinyin tones (MacWhinney, 2017; Pelzl et al., 2020; Wang, 2013). On the other hand, both prominent theories in second language acquisition (SLA) and abundant empirical studies have shown that the ultimate attainment of L2 skills is largely affected by training. An optimized training scheme facilitated with effective instructional programs and tools can greatly, if not completely, offset the age effect in learning a foreign language (Bradlow et al., 1999; Dupoux et al., 2001; Hopp, 2010; Kaan et al., 2007; MacWhinney, 2017; Suzuki & DeKeyser, 2017; Wang et al., 1999). In particular, there has been considerable research showing the effectiveness of various training strategies in facilitating the L2 acquisition of Chinese pronunciation and phonetic knowledge (Li & Dekeyser, 2019; Showalter & Hayes-Harb, 2013; Wang, 2013; Wang et al., 2003).

Despite the significant progress having been made, there are still critical issues underexplored toward a complete understanding of how a foreign tone, or a foreign pronunciation in general, is optimally acquired. The first observation is that most of the previously reported results were based on short-spanned training, constituting an unrealistic replicating of the actual learning. The briefness of the experiment has led to concern about the long-term retention and real-world use of the learned skills, which, after all, is the ultimate purpose of language learning (Li & Dekeyser, 2019; Pelzl et al., 2019). A second observation is that most previous studies were administered in highly restrictive laboratory conditions. While favorable for many reasons, such as confounding factor control, laboratory learning is dramatically different from language learning and use in a daily context, which is far more dynamic and multidimensional. In addition, as noted by Pelzl et al. (2020), for example, most of the previous studies focused only on the L2 suprasegmental learning of a single syllable, providing limited prediction to the learners' tonal identification and production ability at lexical and above levels.

The current study aims to fill these gaps. Accordingly, a longitude experiment, spanning twenty-five weeks from the pretest to the delayed posttest, has been adopted by the current study. Integrated with conventional classroom language instructions that focused on the L2 development of Chinese in general, the current study examines how the phonetic knowledge of multisyllabic words in Chinese could be improved by the online training implemented at the Pinyin Tutor. The design of the Pinyin Tutor was based on a large corpus analysis of the error patterns of the L2 production of Chinese Pinyin, in addition to other innovative features of the platform, such as the in-depth profiling of the learning data and the interactive feedback following each item being trained. Despite the smart design, the Pinyin Tutor needs to be experimentally calibrated in terms of to what extent such intelligent algorithms will be transformed into pedagogical success and what general guidance could be drawn toward an effective smart learning environment in the context of SLA.

Taken together, the experiment reported in the current work investigates how the familiarity level of stimuli will affect the learning outcome of Pinyin for both beginning and intermediate learners of Chinese as an L2 and how such Pinyin learning could be effectively improved through the intelligent CALL environment offered by the Pinyin Tutor. Through a sequence of online Pinyin training and dictation tasks spanned into two semesters in two American universities, we observed an overall significant Pinyin knowledge improvement under either of the training conditions. However, significant

differences were found in the rate of learning gains for different sources of stimuli. It was shown that training with novelty stimuli leads to a greater learning improvement than training with familiar stimuli for intermediate learners. On the other hand, routine training following textbook stimuli leads to a more significant improvement for beginning learners. These findings tend to highlight the importance of embedding the construction of a CALL system with a diversified option of instructional materials and, in particular, underscore how a differentiated feed of novelty stimuli should be catered to different learners at different learning stages so as to efficiently foster their L2 development.

Theoretical framework and literature

It is often argued that adults are generally inferior in learning foreign speech. The critical period hypothesis (CPH) claims that the complete mastery of a foreign phonological system is unattainable for adult L2 learners (Chiswick & Miller, 2008; Lenneberg, 1967; Patkowski, 1990). Indeed, the results of many previous studies and experiments tend to support the belief that the earlier one begins to learn an L2, the better they will pronounce that language (Fathman, 1975; Flege et al., 1995; Moyer, 1999; Oyama, 1982; Patkowski, 1990; Tao et al., 2018; Thompson, 1991; Xiao, 2013, for example). On the other hand, considerable studies have supported the idea that old learners, including adults, are not necessarily inferior in terms of efficiency and levels achievable in L2 learning. For example, Flege (1988) claims that “there is no conclusive support for the existence of a critical period for human speech learning”. Krashen et al. (1979) argue that adults actually learn a foreign language more quickly than children in the early stage, although children may have a better chance of achieving native-like fluency. Hopp (2010) and Donaldson (2011) show that ultimate attainment in L2 is comparable to that of first language (L1), at least in certain aspects of language proficiency with appropriate training. Flege (2018) further emphasizes that input, not age, is the critical predictor of L2 proficiency.

A more comprehensive view from the unified competition model (UCM) is that the ultimate attainment of L2 skills is the outcome of a complex mechanism co-functioned by a variety of protective or risk factors from multiple dimensions. These multidimensional factors constitute the underlay of the age effect and are dynamically alterable by cognitive, behavioral, and environmental interventions (MacWhinney, 1987, 2007, 2014, 2017). Accordingly, adult learners’ development of foreign speech should not be thought of as static. In fact, considerable studies involving a variety of languages have supported the view that auditory perception of language is not completely fixed in adulthood (Bradlow et al., 1999; MacWhinney, 2014; Yan & Sloos, 2019). Some studies indicate that training can lead to improvement in the identification and discrimination of non-native contrasts (Lively et al., 1993; Ingvalson et al., 2012), and even negative phonological transfer can be corrected through repeated training (Crosson et al., 2019; Donaldson, 2011; Flege et al., 1995). With training, non-native speakers of Mandarin Chinese can improve their syllabic tone perception to near-native levels (Sung, 2012; Wang et al., 1999; Wiener et al., 2020). Other studies show that long-term laboratory training has led to generalizable improvement (Bradlow et al., 1999; Suzuki & DeKeyser, 2017).

In particular, the UCM of language learning emphasizes the role of L1 transfer, resonance, and cue validity. In order to promote the acquisition of marked L2 phonemes, adult learners must apply additional learning strategies, including the optimization of high-quality input, selective attention, promotion of L2 resonance, and graduated interval recall (MacWhinney, 2017). According to the classic competition model and the UCM, a highly available and reliable cue will be attached with high strength and advantage in competition (MacWhinney, 1987, 2014, 2017). Therefore, Chinese Pinyin learning is expected to be more robust if valid cues are effectively applied to guide a learner's attention early in language acquisition. In the current study, we focus on the effect of familiarity of stimuli on L2 learning performance, expecting to find out whether familiar or unfamiliar auditory training is more effective in providing optimal contrast and best helping learners to identify, perceive, and acquire the target L2 Chinese speech sound.

There are studies exploring the effect of material familiarity on L2 learning, but mainly in the area of ESL. For instance, Channell (1981) recommends that words be taught in semantically or orthographically homogeneous sets. Neuner (1992) claims that teaching similar words at the same time requires less learning effort. Boers and Lindstrombergh (2005) suggest that lexical chunks of alliterations, such as "jumping and jiggling" or "blue balloon", are easier to learn because of the similar pattern of the components. Nevertheless, recent studies have suggested largely contrary conclusions. Nation (2000) argues that "learning related words at the same time makes learning more difficult. This learning difficulty can be avoided if related words are learned separately". It is recommended that "Teachers can decrease the possibility of interference by making the contexts, collocations, and visual representations of related words as different as possible". Erten and Tekin (2008) show that the learning rate of semantically unrelated words is higher than that of semantically related sets. Similar research studies have been done by Laufer-Dvorkin (2006) and Kim (2016), showing that it is not necessarily effective to teach target words within the context of a mixture of familiar words, at least in the early stages of learning. These controversial conclusions underscore the multidimensionality of SLA hypothesized by the UCM, where the stimulus is co-functioning, in an interwoven manner, with other factors, including scaffolding and chunking, for instance, towards the overall success of L2 learning. The extent to which familiar or novelty words could be facilitative in language learning is ultimately tied to whether they are finely tuned to the L2 learners' developmental levels (MacWhinney, 2014, 2017).

The Pinyin Tutor implementation

Operated and maintained under TalkBank, the Pinyin Tutor is a web-based Mandarin Chinese Pinyin learning and assessment platform programmed with Java. From the learner's end, the Tutor is displayed through the dictation tasks of Chinese Pinyin, covering the pronunciations of all the syntactically possible words or phrases in Chinese. The learners hear the pronunciation of a selected Chinese word by native speakers of Chinese and are asked to enter the Pinyin syllables according to the instructed rules or rules explained in the help menu. Intelligent feedback is provided on different types of errors helping the learner to improve in the following trials. The learners have the option to listen to the correct pronunciation and also to their own attempted pronunciation and compare the subtle differences between the two. Such accurate and tailored feedback is

possible with the Pinyin Tutor because the Tutor has been embedded with the database of all the pronounceable syllables, close to 4000 in total, according to the Chinese Pinyin rules. To date, more than 100 universities or high schools across the world have adopted the Pinyin Tutor in their classroom teaching of Chinese. The following Fig. 1 provides a snapshot of the part of the interface of the Pinyin Tutor, and Fig. 2 demonstrates that the class performance profile can also be readily populated after the completion of each training session. One worthy-to-note feature of the Pinyin Tutor, from a language learning and assessment perspective, is the embedded data analysis function programmed in Java. Learners' L1 backgrounds and past L2 learning experiences are pooled upon registration. In addition, learning portfolios, such as the time of login and logout, the duration of the drill, the error type, the session-wise score, score breakdown, and the number of attempts on each item, are automatically generated on a real-time basis upon completion of each training. In addition to helping learners or class instructors to optimize the learning ladder, such a function also provides an essential empirical basis for educators and researchers to calibrate and improve the existing pedagogical theories and practice.

The experiment of the current study trained two groups of learners learning Chinese phonetic knowledge through the Pinyin Tutor. One group of learners was trained with familiar words from the textbook and the other with novel words that are outside of the textbook. The design of familiar and novel cues reflected the different roles they play in phonetic knowledge development among L2 learners. By replicating a stimulus with which a past activation was produced in the auditory cortex, a familiar word serves as an anchor for the learner to recall a specific Pinyin pronunciation and spelling with less cognitive load. However, novel words are expected to promote cue generality by allowing learners to pick up more comprehensive knowledge that can be applied to new forms. The rule of minimum pair was applied by the current study to design the novelty stimuli,

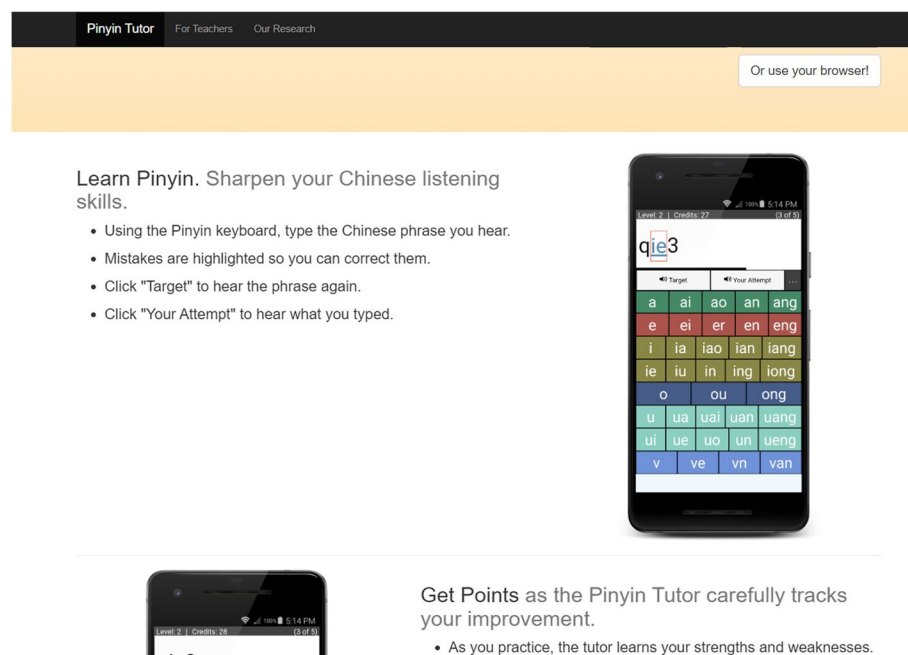


Fig. 1 Snapshot of part of the interface of the Pinyin Tutor

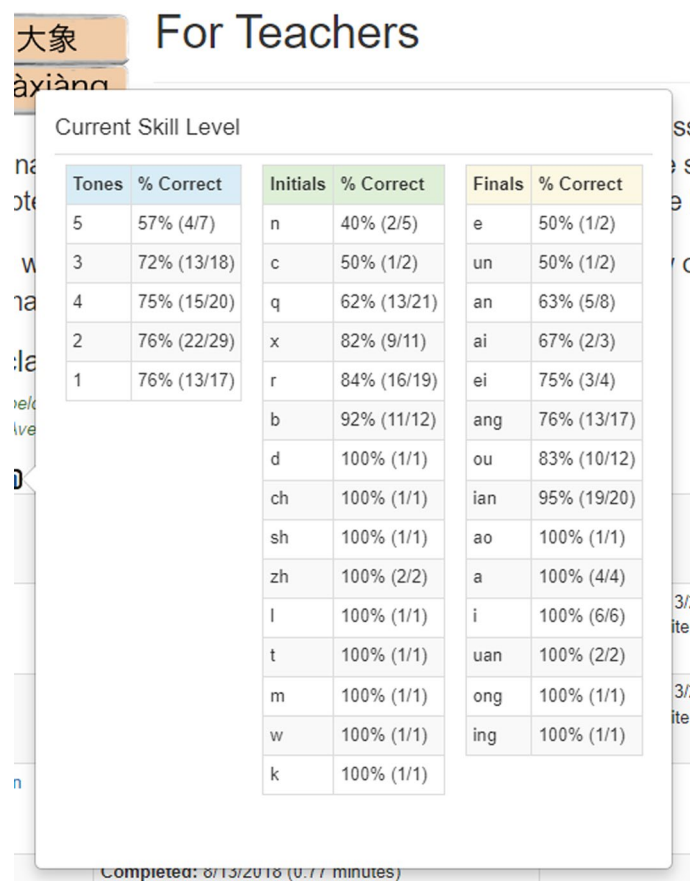


Fig. 2 Example of the class performance profile after a training at the Pinyin Tutor

which consisted of new words that were structurally matched with words from the textbook in phonetic spelling but with different meanings. For example, to match a textbook word “jie4shao4” (to introduce), a novel word “jie1shou4” (to accept) not appearing in the textbook was selected as the counterpart novel word. One worthy note is that all novel stimuli designed were valid words in use in the contemporary Chinese language. They were prompted to the learners with minimum variations, either in initial or final, in comparison to those familiar words presented in the textbook so as to stimulate the learner’s generalized knowledge of Pinyin. From the UCM perspective, resonant practice with familiar words is expected to be more suitable for novice learners to strengthen their basic skills in auditory phonology, whereas novel words are more useful for more advanced learners, whose metalinguistic awareness of the L2 will make such a generalization more probable.

Method and procedure

Beginning learners of Chinese from Carnegie Mellon University (CMU) and intermediate learners from Pennsylvania State University (PSU) were invited to the experiment. Because of the *in vivo* nature of the study, not all invited students completed all the Pinyin Tutor sessions. A student must have completed at least two-thirds of the offered

training sessions and attended both the pretest and the posttest for the corresponding learning data to enter into the analysis. In addition, students with pretest score in Pinyin knowledge above 90 were also not included in the subsequent analysis to minimize the ceiling effect of test. According to these thresholds, there were effectively 58 beginning learners of Chinese from CMU and 36 intermediate learners from PSU who participated in the current study. At the end of the training, students were asked to fill in a questionnaire surveying their language background, self-rated learning interest and diligence, and evaluations of the Pinyin Tutor. From the survey, we have that the beginning learners' L1 backgrounds are English (41%), Korean (37%), Cantonese heritage language (5%), Mandarin heritage language (5%), and others (12%). The L1 backgrounds of the intermediate learners are English (60%), Korean (16%), Cantonese heritage language (5%), Mandarin heritage language (14%), and others (5%).

The allocation of participants at CMU was randomized according to the students' campus ID names provided to the Pinyin Tutor coordinator. Depending on the total number of letters in the ID string being odd or even, a participant was assigned to one of the two training conditions: the routine condition in which the stimuli were familiar words from the textbook they were using and the corresponding lesson had been covered in their Chinese class at the time of training or the novelty condition in which the training stimuli were not the words in the textbook but were designed according to the rule of minimum pair in comparison to words appeared in the textbook. At PSU, the allocation of the participants was randomized in accordance with the label of the class in which a student was enrolled. A randomly selected three classes of students were assigned to the textbook or routine condition, and students from the other three classes were assigned to the non-textbook or novelty condition. Class-wise pre-training proficiencies of the intermediate participants did not exhibit statistical difference with $p\text{-value} > 0.70$ for the F-test for comparison of means of the pretest scores of the six classes.

All the online training sessions were conducted at the Pinyin Tutor at the TalkBank (<http://talkbank.org/Pinyin/>). A preliminary session was arranged before the pretest focusing on the tutorial on how to use the Pinyin Tutor. In particular, the following rules for entering Pinyin in the Pinyin Tutor were instructed and familiarized by the students:

1. Use numbers (1, 2, 3, 4) for the four tones. For example, "tóngxué" would be written as "tong2xue2".
2. Use number 5 for the neutral tone. For example, "shénme" would be written as "shen2me5".
3. Use "v" for the umlauted "ü", as in "nv3er2" for "nǚér".
4. Punctuations (such as apostrophes or hyphens) are not accepted by the Tutor.

In the training, a learner is prompted by an audio recording of a word in standard Chinese, after which they must correctly enter the Pinyin of the word. If the answer is correct, the student progresses to the next trial; if incorrect, they are given a second chance, along with feedback regarding the difference between their form and the target. If they fail on the second attempt, the Tutor remembers the incorrect word and presents it to the learner later. On each trial, students also have the option of comparing the sound of

the Pinyin that they typed with the target sound following each of the two Pinyin entry attempts.

A pretest was administered at the beginning of the experiment for both beginning learners and intermediate learners. The beginning learners received two weeks of conventional classroom instruction and learning focusing on Pinyin knowledge before the pretest was arranged in the following week, right after the Pinyin topic was covered. The intermediate learners received a similar classroom introduction of Pinyin one or two semesters ago and were provided with a class review of Pinyin lasting about 40 min before the pretest was assigned to them in the following week. The pretest was composed of 40 novel words, including eight monosyllabic words, 28 disyllabic words, and four multisyllabic words. A posttest was assigned in the thirteenth week of the experiment. A delayed posttest was conducted twelve weeks after the posttest. The posttest and the delayed posttest used the same Pinyin items contained in the pretest but in re-randomized order. Students were only instructed to complete the Pinyin Tutor sessions as part of the take-home assignments of the course they were taking. They were not informed of the posttest or the delayed posttest after the training. Due to practical constraints, a delayed posttest was not administered for the intermediate learners.

Six training sessions were assigned biweekly as taking-home Pinyin exercises through the Pinyin Tutor. Each training session contains 20–30 Pinyin items, depending on the contents and vocabulary covered in the corresponding classroom instruction. Words selected from the textbook were presented to students in the same order as they appeared in the textbook vocabulary list. The novelty stimuli were chosen from a wide pool of general words and were structurally matched with words from the textbook according to the minimum pair rule of design. For example, a novelty counterpart of “xue2xi2” (to study) in the textbook could be “xue2qi1” (semester) or “que1xi2” (being absent). Therefore, phonologically, the difficulties of words used in the two treatment groups were equivalent, while semantically, the words in the textbook group were more familiar to the learners.

A Perl parser partitioned all the training stimuli and the test items into initials, finals, and tones to examine the accuracy rate and the error pattern for in-depth analysis. In order to make sure that the confusion was not caused by the quality of the recording or other technical errors, two native speakers were invited to evaluate the training stimuli of the experiment. The testers accessed the online Pinyin Tutor and went through the Pinyin training sessions exactly as the learners did. The testers’ answers were parsed in the same way as if they were student attempts. The mismatches between targets and attempts of the native speakers were sorted out. These mismatches, accounted for 1.5% of the total stimuli tried, were marked as special scenarios for further study, the corresponding items of which were not included in the subsequent analysis of the current study.

Results and analysis

The first research question concerns which of the two training modes, training with familiar or routine stimuli, is more effective for beginning learners of Chinese as an L2. As indicated by the following Fig. 3 for pretest–posttest comparison (sample size $N=28$ for routine and $N=30$ for novelty) and Fig. 4 for pretest–delayed posttest

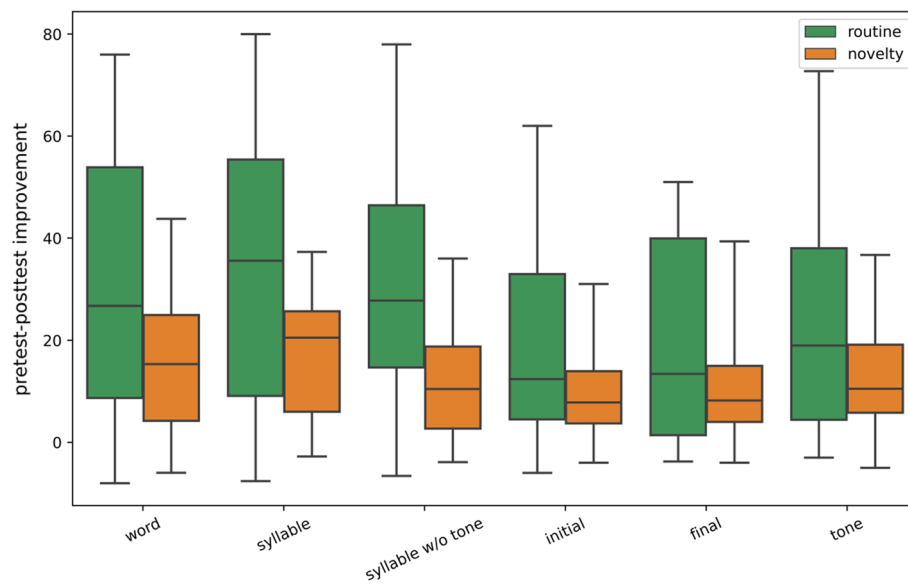


Fig. 3 Beginners' pretest–posttest improvements under the two training conditions

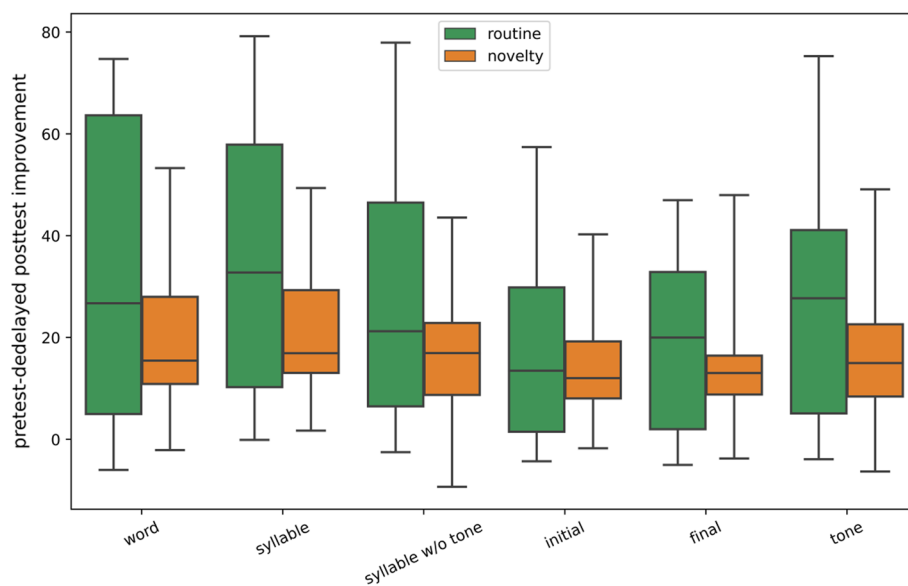


Fig. 4 Beginners' pretest-delayed posttest improvements under the two training conditions

comparison ($N=25$ for routine and $N=16$ for novelty), training with familiar stimuli leads to a greater improvement among beginning learners, on average, in every component of Pinyin knowledge, although the increments in word and syllable scores are typically substantially higher than those in initial and final. More specifically, as shown by Table 1, the mean pretest–posttest improvement of the routine condition is statistically higher than that of the novelty condition in each component of Pinyin knowledge with p -value < 0.05 except in the initial component with p -value slightly higher than 0.10. This relatively less significant difference in the performance of the initial component under

Table 1 Pretest–posttest improvements among beginning learners

	Routine		Novelty		t-test	
	Mean	SD	Mean	SD	t-statistic	p-value
Word	32.14	27.71	19.27	14.58	2.2349	0.0147
Syllable	33.92	26.71	19.73	13.44	2.5817	0.0062
Syllable w/o tone	29.76	21.37	15.99	11.14	3.1071	0.0015
Initial	17.81	17.33	13.29	8.40	1.2776	0.1033
Final	19.18	19.87	11.20	9.62	1.9674	0.0271
Tone	23.55	23.31	14.52	10.44	1.9258	0.0296

Table 2 Pretest–delayed posttest improvements among beginning learners

	Routine		Novelty		t-test	
	Mean	SD	Mean	SD	t-statistic	p-value
Word	33.12	29.75	19.56	15.09	1.6851	0.0500
Syllable	34.33	26.9	20.83	13.92	1.8489	0.0360
Syllable w/o tone	29.02	25.67	17.09	13.59	1.7067	0.0479
Initial	18.25	19.57	13.62	10.03	0.8727	0.1941
Final	17.48	16.89	13.43	11.47	0.8406	0.2029
Tone	24.83	22.36	16.12	13.42	1.4021	0.0844

the two conditions is primarily an indication that the sub-lexical segmental perception of Pinyin is relatively less challenging for L2 learners to acquire, although such acquisition in itself may not be strong enough to predict the speaking proficiency of Chinese in a natural conversational context. Such a disparity in learning challenge between initial and word for novice learners was statistically significant at a 95% confidence level with $p\text{-value} < 0.01$ for the t-test on the pretest scores of these two components of Pinyin. When it comes to the pretest–delayed posttest comparison, as shown by Table 2, the advantage of routine training has dwindled so that the difference between the two training conditions is statistically significant, at a 95% confidence level, only in word, syllable, and syllable without tone, whereas insignificant in initial, final, and tone. Again, such subtle contrast in terms of at which component of Pinyin the routine training is more strongly effective for beginning learners partially reflects the challenge to synthesize the segmental and suprasegmental skills into the knowledge demanded at a higher framework of language proficiency (MacWhinney, 2014).

As revealed by Tables 1 and 2, the improvement from the pretest to the delayed posttest is greater than that from the pretest to the posttest for both training modes, demonstrating that favorable retention is achieved from the Pinyin Tutor training. This is in spite of the fact that such a difference is rather subtle and does not yet constitute a statistical significance with $p\text{-value} = 0.1647$ for testing the null hypothesis that the improvement from pretest to delayed posttest is higher than that from pretest to posttest. While an accurate account of the mechanism of retention is not trivial, the observed strong performance in the delayed posttest appears to be well in line with the lag effect of learning (Kahana & Howard, 2005). For instance, approximately eight days of interval for recall was demonstrated by Kapler et al. (2015) as optimal for long retention of

natural science knowledge. By providing a relatively more natural learning and recalling environment, the approximately biweekly spaced and repeated Pinyin Tutor training of the current study, spanning a semester-long period in total, exhibits the potential to propagate the retention into the long future after training. Nevertheless, we leave the post hoc explanation of the observed retention open to the confounding of other factors, including the motivation of the beginning learners, as many of this group of learners would continue to learn Chinese in subsequent semesters. In addition, there is also one exception to the above observation: the pretest-delayed posttest improvement in final is lower than that of the pretest–posttest by 8.9% with $p\text{-value}=0.3920$ for testing the null hypothesis that the two improvements are equal. A separate study of error analysis based on the whole learners' corpus generated by the Pinyin Tutor shows that diphthongs and nasal vowels pose persistent perceptual difficulties even to advanced L2 learners of Chinese with an L1 background in Korean, for example. In particular, for learners with L1 in Korean, the overall confusion rate in the Pinyin identification tasks of the nasal vowel with glide, /iong/, is as high as 56%, compared to a low confusion rate of 5% in the identification of /b/, the unaspirated dental initial in Pinyin. Overcoming such persistent challenges entails the support of more advanced and focused training schemes (MacWhinney, 2014).

It is also observed that learners under the textbook training condition universally exhibited larger variances in their learning improvement compared to their counterpart. With other conditions equal, routine training with familiar stimuli from textbook entails the ability to infer the Pinyin rules from a limited, confined context. Benefits of such training, from the UCM perspective, include efficient auditory access and processing, easy formation of strong chunkings, and positive metalinguistic resonance and transfer when the familiar stimuli effectively subserve a high cue validity. On the other hand, familiar stimuli may also give rise to risk factors such as negative transfer and entrenchment in the auditory activation. Such polarized scenarios are inductive to the highly varied learning improvements among beginning learners. The results show that novelty stimuli, with their wide-sourced nature, have helped to smooth out such odds in the current study.

The second research question hypothesizes that training with novel words will be more effective than familiar words for intermediate learners. As demonstrated by Fig. 5, the pretest–posttest improvement of the group under novelty stimuli condition ($N=15$) is significantly greater than that under familiar stimuli condition ($N=21$) in every component of Pinyin knowledge. More specifically, as Table 3 shows, the learning-enhancing advantage of novelty condition over routine condition is significant at 99% confidence level ($p\text{-value}<0.01$) in syllable, syllable without tone, and initial dimensions of Pinyin knowledge, while such an advantage is significant at 95% confidence level ($p\text{-value}<0.05$) in final, tone, and word dimensions. The results are overall consistent with the pedagogical implications of the UCM, where the multifaceted dynamical nature of SLA necessitates a graduated reformation of learning and teaching strategy finely tuned to the learners' developmental level of linguistic and metalinguistic skills. In particular, proportionally increasing the input of unfamiliar stimuli and exposure to more contextualized circumstances should be inducing to effective chunking, positive transfer, and efficient cortex activation. According to the curriculum and the learning experience of the

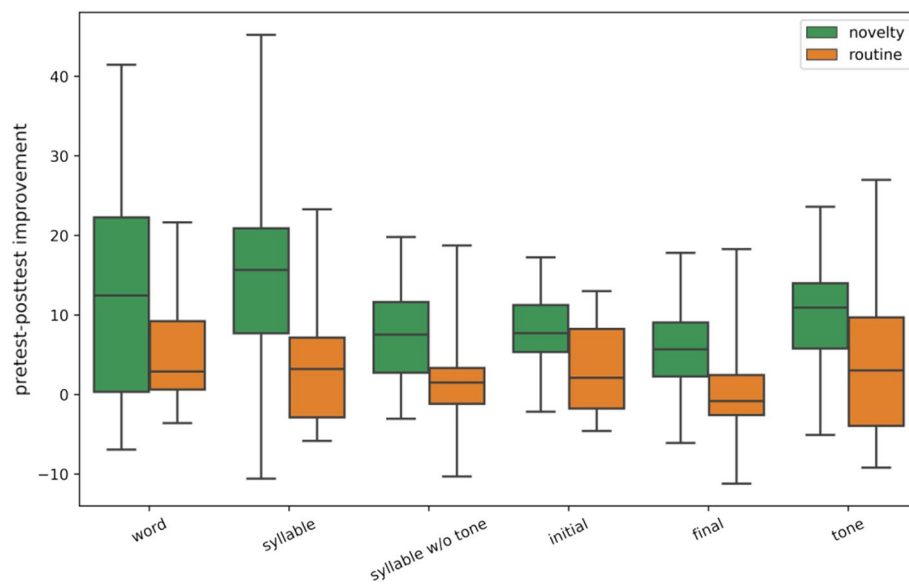


Fig. 5 Intermediate learners' pretest-delayed posttest improvements under the two training conditions

Table 3 Pretest–posttest improvements among intermediate learners

	Routine		Novelty		t-test	
	Mean Pre-Posttest Increase	SD	Mean Pre-Posttest Increase	SD	t-statistic	p-value
Word	5.40	6.85	12.59	13.05	1.9442	0.0301
Syllable	3.36	7.64	14.74	12.94	3.0421	0.0023
Syllable w/o tone	1.71	7.20	7.21	6.09	2.4760	0.0092
Initial	3.50	5.729	8.20	4.87	2.6554	0.0060
Final	1.21	6.38	6.14	6.07	2.3639	0.0120
Tone	4.31	10.80	10.02	7.01	1.9279	0.0311

intermediate learners, by the time of experiment, they should have learned the Pinyin transcription system at least one semester before and have developed basic communicative skills in all of the four aspects of the Chinese language, namely, listening, speaking, reading, and writing. They supposedly have acquired about 150 to 300 lexical units of Chinese, meeting the basic conversational needs in the daily scenes such as greeting, dining, socializing with friends, and traveling. Thus it is not surprising that training with novelty stimuli has more robustly sharpened these learners' perceptual sensitivity to Pinyin and hence their dictation performance in the phonetic identification tasks. Put broadly, the comparative benefit of novelty training to intermediate learners in the current study is a manifestation of their generalized metacognitive readiness to encode and process more diversified L2 information at lexical or above level.

Notwithstanding, as shown in Table 3 and Fig. 5, the pretest–posttest improvements among intermediate learners are generally smaller than those among beginning learners. This gap is firstly an indication of the floor effect of test as intermediate learners had learned Pinyin before and had stronger metalinguistic abilities to rely on for the Pinyin

identification tasks in the pretest. However, the beginning and intermediate learners were from different universities, implying the differences in the *in vivo* instructional style as well as the hardware supporting the Pinyin Tutor, which may facilitate or impede the learning experience. It could also give rise to differentiations in the L2 course entry conditions and, correspondingly, the differentiations in learners' backgrounds and learning motivation. Thus, the influence of such potential confounding factors should not be ruled out when it comes to interpreting such a gap in the learning gain between the two groups of learners.

Similar to that for beginning learners, an observation to highlight is that learners under the routine condition generally demonstrated more significant variances in learning improvement relative to the learners under the novelty condition. However, such disparity in variances is not as apparent as that for beginning learners, and there are also exceptions to such observations in initial and tone. As discussed, training with familiar words posits not only high benefits, such as efficiency in cue activation and auditory processing, but also a high likelihood of entrenchment. The results tend to evince that such a polarizing effect on learning performance variances could be mitigated by training with novel words from more diversified and more complete lexical sources.

An overall evaluation of the learning-enhancing effect of the Pinyin Tutor can be achieved by combining the pretest–posttest results for both the routine and novelty training conditions. For beginning learners, as shown in Fig. 6, the consolidated percentage pretest–posttest improvements are 67%, 52%, 36%, 21%, 21%, and 28%, respectively, for word, syllable, syllable without tone, initial, final, and tone. And the corresponding percentage improvements for pretest–delayed posttest improvements are 73%, 52%, 38%, 21%, 22%, and 31%, in the same order of the respective components of Pinyin, demonstrating stronger long-term retention in comparison to the short-term acquisition. The high retention of the syllabic level of Pinyin knowledge in the current study is consistent, in principle, with a couple of previous studies regarding the acquisition and production of foreign speech sounds, including Wang et al. (2003) and Li and Dekeyser (2019) for

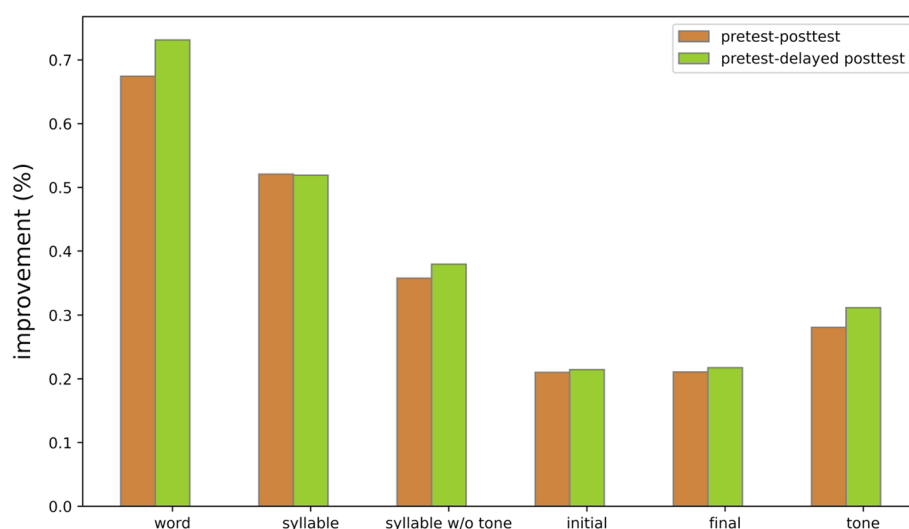


Fig. 6 Overall learning improvement among beginning learners through the Pinyin Tutor

the perception and identification of Pinyin tones. These findings, put together, underline the critical role of training in SLA in general. More specifically, the UCM postulates the relative plasticity of the auditory cortex, allowing adult L2 learners to retune the underlying mental representation toward foreign phonetic perception and production, where the finely tuned input and training, along with scaffolding the individual differences, is the key inducing factor for such auditory plasticity to be exploited. For intermediate learners, the overall effectiveness of the online training by the Pinyin Tutor is also highlight-worthy in that the respective consolidated percentage pretest–posttest improvements are 15%, 12%, 5%, 6%, 4%, and 8% for the six components of Pinyin under consideration, although these increments are substantially lower than those for beginning learners primarily due to the floor effect of test among other potential confounding factors.

Discussion

Two experiment conditions focusing on the types of word stimuli for the Pinyin Tutor training were administered by the current study, with the awareness that an arrangement of another randomized control group not using the Pinyin Tutor throughout the semester would provide a more comprehensive picture regarding to what extent the Pinyin Tutor is advantageous in comparison to other conventional learning modes. Such a third design was not adopted due to the *in vivo* nature of the experiment, particularly in that all students who participated in the training were motivated not only by to enhance their Chinese skills but also to earn course credit, thus expecting fairness in learning support and resources. Nevertheless, a reckoning of the learning-enhancing power of the Pinyin Tutor is still possible by comparison with the existing results on L2 phonetic knowledge acquisition. For instance, previous research suggested that a two-week focused training in lab conditions helped CFL learners to better perceive the tonal information of Chinese by a 21% increase in accuracy as tested through Pinyin identification tasks (Wang et al., 1999). Also, an average of 18% improvement in the Pinyin tone production among American CFL learners was reported by Wang et al. (2003) after a two-week lab training. With a training length restricted to within one week, Wayland and Li (2008) demonstrated that L2 learners of Thai improved, largely by less than 10%, their perception of the tonal contrast in Thai. In the current study, the phonetic knowledge of Pinyin at tone level improved by 31% in long-term retention among beginning learners of Chinese through the training offered by the Pinyin Tutor. Moreover, the performance increment in the word or syllable level is shown as even higher. Taken together, the Pinyin improvement achieved by the participants in the current study, particularly those at the beginner level, appears plausible enough to endorse the effectiveness of the Pinyin Tutor. This said, all such comparisons should be taken with caution. First, the training conditions were diversified in various aspects, including contents, duration, and intensity. More importantly, one should be aware of the long-term difficulties in the perception and production of a foreign speech, especially in real conversational contexts.

In the current study, the same set of Pinyin items was used, with the sequential order randomized, for the pretest and posttest. The main benefit of such a design is the consistency of the level of difficulty of the tests. The nature of the tasks in the current study, i.e., to transcribe the Pinyin into its phonetic forms with tonal identifications instantly

after listening to their pronunciations in a CALL environment, should warrant the comparability of test difficulty between the pretest and delayed posttest. One possible concern is that one participant might have simply mechanically memorized the test items in the pretest. Such a scenario is not likely the case in reality. First, the test items were prompted to the participants in a very short time with randomized order so that it was not easy to memorize them in such a brief acquaintance. Second, past studies show that human memory, especially about context-void speech stimuli, is extremely brief and may decay as quickly as in hours (Higgins et al., 2014). Furthermore, the posttest or delayed posttest was not announced before or during the pretest, so the participants had no motivation to memorize the test items. In addition, given the intensity of the course and the typical high academic load of other subjects for a typical American university, it is not likely that they sought additional venues and training programs, other than those provided by the in vivo exercises, to learn Pinyin throughout the experiment, which had also been qualitatively confirmed by the collected post-experiment survey. So it is believed that the history factor should be confidently ruled out for the analysis.

In summary of the findings of the current study, guidance for constructing a smart learning environment should start with the identification of the learning and instructional objectives that are both justifiable from subject theories and achievable through pedagogical practice. In the setting of SLA, as concerned by the current study, the incremental input of novelty stimuli with optimal scheduling is demonstrated as a supportive instructional strategy towards enhanced learning. Another essential tip for smart learning is to meticulously address the learning needs with individual differences taken into account, which may include tailored task design and interactive feedback, for instance. Learners' profiling and learning data analysis are almost sine qua non for nowadays smart learning environments (see also Zhu et al., 2016). These said, it is not our stand to underestimate the multifacetedness and the challenges of designing, constructing, and updating any smart learning environment at the practical level. Instead, a robust and adaptable smart learning surroundings should allow for differentiating the overall learning goal by focusing on a particular aspect of knowledge or a specific learning strategy, such as the acquisition of phonetic knowledge of Pinyin using novelty stimuli as one selective training option offered by the Pinyin Tutor in the current study.

Conclusion and limitations

The current study concerns the effect of the familiarity level of training stimuli on the acquisition of Pinyin knowledge for L2 learners of Chinese. Two training conditions, i.e., training with familiar stimuli from textbook and unfamiliar stimuli from novelty design, were administered to two groups of learners in terms of their initial L2 level. The results show that training with familiar stimuli is more effective in facilitating the beginner's L2 development of Pinyin, whereas training with novelty stimuli is more effective for intermediate learners. The results underscore the critical role of differentiated task design and graduated input for optimally fostering the SLA as learners' metalinguistic ability advances, as implied by the UCM. As shown, an approximately biweekly spaced self-monitored online training spanning throughout a normal semester, with novel stimuli designed pairing to the classroom-taught vocabulary, has proven to be of significant

learning fostering benefit for the L2 acquisition of Pinyin knowledge at all dimensions. These findings shed enunciating light on real pedagogical designs in terms of optimizing the instructional curriculum and learning strategies.

The training in the current study was offered by the Pinyin Tutor at TalkBank. Consolidating the results for the two treatment groups, i.e., the group under familiar stimuli and the other under novelty stimuli, the learning-enhancing effectiveness of the Pinyin Tutor has been confirmed. Particularly strong benefits have been observed among beginning learners, with at least 21% pretest–posttest performance increment in all the component Pinyin skills under assessment. More importantly, equivalent or stronger retentions of the acquired knowledge at all aspects have been proved by the delayed posttest. The results are, in principle, consistent with or statistically more substantial than the considerable previous studies regarding the perception and identification tasks of L2 speech sound, e.g., Wang (2013), Wang et al. (2003), and Kaan et al. (2007). Given that the study was based on the Pinyin learning at the lexical and above level and was administered in a semester-long natural learning environment instead of restrictive lab conditions, the results are also expected to shed fresh insights into the debate casted by studies such as Pelzl (2019) where a relatively less confident picture in L2 acquisition of a tonal language was suggested.

As reminded by MacWhinney (2017) and Yang et al. (2018), for instance, technology in itself is not sufficient to benefit learning. Overall, the satisfactory performance of the Pinyin Tutor underscores the importance of coherent integration of language learning theories in the CALL design. One critical aspect of such design, as proved by the current study, centers around optimizing the learning input and scheduling. For specific, differentiated learning tasks with a finely tuned increment of novelty stimuli, coupled with a graduated recall with a timeframe of around two weeks, have proven significantly effective in fostering the acquisition of the lexical level of Pinyin knowledge as adult L2 learners progress from novice to higher levels. The Pinyin Tutor has also constituted a plausible example where a successful CALL platform should be able to effectively integrate with traditional classroom instructions, instead of seeking to completely replace them, so that the benefit of face-to-face conversation with instructors is fully leveraged. Broadly defined as an interactive media, the Pinyin Tutor is also flexible and robust to integrate with other emerging learning modes, including situated learning and blended and flipped instruction, that have been proven effective in enhancing knowledge acquisition in general (Tong et al., 2020; Wei et al., 2020; Zhu et al., 2016).

One limitation of such in vivo study is the difficulty of test administration and training control. The large noise magnitude may partly account for the less significance of some statistical parameters and hypotheses, such as the exceptionally high p-value of 0.1033 for the pretest–posttest improvement comparison between the routine and novelty training groups. A natural direction for future study is to administer similar experiments with more uniform training conditions and fewer noises at various aspects so that the educational performance of the Pinyin Tutor could be more fully evaluated on various fronts. With a total number of participants of 58 throughout the pretest–posttest experiment, the sample size, although worked finely for ANOVA, is still limited when a more rigorous quest is desired in terms of covariate structures of the large-number dimensions of the L2 knowledge or prediction-focused modeling and validation. In addition,

future research should endeavor to include more types of learning factors as premised by the UCM for completeness. Future upgrading of the Pinyin Tutor in hardware and algorithmic construct is also commendable, including incorporating more speech processing functions and developing mobile-friendly Apps.

Abbreviations

L2	Second language
CALL	Computer-assisted language learning
SLA	Second language acquisition
CPH	Critical period hypothesis
L1	First language
CMU	Carnegie Mellon University
PSU	Pennsylvania State University
UCM	Unified competition model

Acknowledgements

The authors would like to thank all the students and teachers who participated in the training experiment for the current study. The authors are grateful for the constructive and valuable comments from the anonymous referees for improving the earlier version of the paper.

Author contributions

The authors jointly formulated the study, designed the experiment and procedure, and jointly collected the data and conducted preliminary analysis. YZ surveyed the literature, analyzed the data, carried out the empirical analysis, and drafted the initial manuscript. BM provided critical comments on the UCM and the results. Both authors read and approved the final manuscript.

Funding

There is no funding specifically received for the current study.

Data availability

Data used for the current study are available upon reasonable request for academic researches.

Declarations

Informed consent

Informed consent was obtained from all the participants involved in the study.

Competing interests

The authors declare no competing interests.

Received: 26 September 2022 Accepted: 4 January 2023

Published online: 10 January 2023

References

- Boers, F. and Lindstrombergh, S. (2005). Finding ways to make phrase-learning feasible: the mnemonic effect of alliteration. *System* 33, 225–238.
- 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.
- Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal*, 35, 115–122.
- Chiswick, B. R., & Miller, P. W. (2008). A test of the critical period hypothesis for language learning. *Journal of Multilingual and Multicultural Development*, 29(1), 16–29.
- Crosson, A. C., McKeown, M. G., Moore, D. W., & Ye, F. (2019). Extending the bounds of morphology instruction: Teaching Latin roots facilitates academic word learning for English Learner adolescents. *Reading and Writing*, 32(3), 689–727.
- Donaldson, B. (2011). Left dislocation in near native French. *Studies in Second Language Acquisition*, 33(3), 399–432.
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress “deafness.” *The Journal of the Acoustical Society of America*, 1103(Pt 1), 1606–1618.
- Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36, 407–422.
- Fathman, A. (1975). The relationship between age and second language productive ability. *Language Learning*, 25, 245–253.
- Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism Language and Cognition*, 21, 919–992.
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages. *Applied Linguistics*, 8, 162–177.
- Flege, J. E. (1988). Using visual information to train foreign language vowel production. *Language Learning*, 38, 365–407.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.

- Higgins, D., Graupner, M., & Brunel, N.J. (2014). Memory maintenance in synapses with calcium-based plasticity in the presence of background activity. *PLoS Computational Biology*, 10.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120(4), 901–931.
- Ingvalson, E., Holt, L., & McClelland, J. (2012). Can native Japanese listeners learn to differentiate /r-1/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, 15, 255–274.
- Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2007). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience*, 9, 53–53.
- Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review*, 12, 159–164.
- Kapler, I., Weston, T., & Wieseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, 36, 38–45.
- Kim, C. (2016). L2 learners' recognition of unfamiliar idioms composed of familiar words. *Language Awareness*, 25(1–2), 89–109. <https://doi.org/10.1080/09658416.2015.1122025>
- Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly*, 13, 573–582. <https://doi.org/10.2307/3586451>
- Laufer-Dvorkin, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review*, 63(1), 149–166.
- Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.
- Li, M., & DeKeyser, R. M. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *The Modern Language Journal*, 103(3), 607–628.
- Lively, S., Logan, J., & 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 Pt 1), 1242–1255.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2007). A unified model. In N. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition*. Oxford: Lawrence Erlbaum Press.
- MacWhinney, B. (2017). A unified model of first and second language learning. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Sources of variation in first language acquisition: Languages, contexts, and learners*. pp. 287–310.
- MacWhinney, B. (2014). What we have learned. *Journal of Child Language*, 41, 124–131. <https://doi.org/10.1017/S0305000914000142>
- MacWhinney, B. (2021). The competition model: Past and future. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A life in cognition: studies in cognitive science in Honor of Csaba Pléh* (pp. 3–16). Springer.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: the critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21, 81–108.
- Nation, P. (2000). Learning vocabulary in lexical sets: dangers and guidelines. *TESOL Journal*, 9(2), 6–10.
- Neuner, G. (1992). The role of experience in a content- and comprehension-oriented approach to learning a foreign language. In P. J. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 156–166). MacMillan.
- Oyama, S. (1982). The sensitive period for the acquisition of a nonnative phonological system. In: S. Krashen, R. Scarcella, & M. Long (Eds.), *Child-adult differences in second language acquisition*. Rowley, Newbury.
- Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11, 73–89.
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. M. (2020). Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin. *Studies in Second Language Acquisition*, 43, 268–296.
- Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, 29(2), 185–200.
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, 53, 273–293.
- Sung, K. (2012). A study on Chinese-character learning strategies and character learning performance among American learners of Chinese. *Chinese as a Second Language Research*, 1(2), 193–210.
- Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research*, 21(2), 166–188.
- Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, 123(1), 397–413.
- Tao, H., Salaberry, M., Yeh, M., & Burch, A. (2018). Using authentic spoken language across all levels of language teaching: Developing discourse and interactional competence. *Chinese as a Second Language Research*, 7(1), 1–13. <https://doi.org/10.1515/caslar-2018-0001>
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41, 177–204.
- Tong, Y., Kinshuk, & Wei, X. (2020). Teaching design and practice of a project-based blended learning model. *International Journal of Mobile and Blended Learning*, 12, 33–50.
- Wang, X. (2012). Auditory and visual training on mandarin tones: A pilot study on phrases and sentences. *International Journal of Computer-Assisted Language Learning and Teaching*, 2, 16–29.
- Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, 97, 144–160.
- 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.
- Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36, 250–267.

- Wei, X., Cheng, I., Chen, N., Yang, X., Liu, Y., Dong, Y., Zhai, X., & Kinshuk. (2020). Effect of the flipped classroom on the mathematics performance of middle school students. *Educational Technology Research and Development*, 68, 1461–1484.
- Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' mandarin tone productions? *Modern Language Journal*, 104(1), 152–168.
- Xiao, Y. (2013). Effect of home background on advanced heritage language learning. *Chinese as a Second Language Research*, 2(2), 193–220. <https://doi.org/10.1515/caslar-2013-0030>
- Yan, M., & Sloos, M. (2019). Perceptual cue weighting in L1 and L2 Chinese: An investigation of aspiration and frication in Chinese obstruents by Danish learners. *Chinese as a Second Language Research*, 8(1), 57–83. <https://doi.org/10.1515/caslar-2019-0003>
- Yang, J., Pan, H. W., Zhou, W., & Huang, R. (2018). Evaluation of smart classroom from the perspective of infusing technology into pedagogy. *Smart Learning Environments*, 5, 1–11.
- Zhang, Y. (2016). The role of morphological awareness in the incidental learning of Chinese characters among CSL learners. *Language Awareness*, 25(3), 179–196.
- Zhu, Z., Yu, M., & Riezebos, P. (2016). A research framework of smart education. *Smart Learning Environments*, 3, 1–17.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
