LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Learning a Tonal Language by Attending to the Tone: An In Vivo Experiment

Ying Liu
University of Pittsburgh and Liaoning Normal University

Min Wang
University of Maryland

Charles A. Perfetti
University of Pittsburgh

Brian Brubaker
Carnegie Mellon University

Sumei Wu
Carnegie Mellon University

Brian MacWhinney
Carnegie Mellon University

Learning the Chinese tone system is a major challenge to students of Chinese as a second or foreign language. Part of the problem is that the spoken Chinese syllable presents a complex perceptual input that overlaps tone with segments. This complexity can be addressed through directing attention to the critical features of a component (tone in this case) within a complex perceptual input stimulus. We tested hypotheses based on this feature-focusing assumption in an in vivo classroom setting. First-year students in a Chinese language program at a U.S. university were trained to identify the

This study was supported by the Pittsburgh Learning Science Center funded by National Science Foundation award SBE-0354420. We thank all the students in the Chinese Program at Carnegie Mellon University for their active participation. We are also grateful to Dayne Grove for his assistance in programming the online tutor.

Correspondence concerning this article should be addressed to Min Wang, Department of Human Development, Institute for Child Study, 3304 Benjamin Building (#143), University of Maryland, College Park, MD 20742. Internet: minwang@umd.edu

DOI: 10.1111/j.1467-9922.2011.00673.x

tones of 228 syllables learned across eight lessons in the first semester. Three learning conditions were designed to support tone learning by presenting (a) visual pitch contours that depict the acoustic shape of the tones, together with pinyin spelling of the spoken syllables (Contour + Pinyin condition); (b) numbers that represent the tones in traditional computer interface, together with pinyin spelling of the spoken syllables (Number + Pinyin condition); and (c) visual pitch contours without pinyin spelling (Contour Only condition). Analyses of student activity logs (learning curves) and pretests and posttests showed significant effects of learning condition. The results suggested that the Contour + Pinyin condition had more error reduction in tone recognition over the activity log than the Contour Only condition and greater improvement from the pretest to posttest than the Number + Pinyin condition. These findings point at the value of separate support for the two major components (tone and segments) of a tonal language.

Keywords Chinese as a second/foreign language; spoken syllable; tone; pinyin; pitch contour; in vivo experiment; learning curves; online instruction

Introduction

The tonal feature of spoken Chinese poses a particular challenge for a beginning learner in speaking and reading Chinese (Wang, Perfetti, & Liu, 2003). At a practical level, this difficulty presents a challenge to teachers and learners of Chinese. At a theoretical level, it addresses two contrasting ideas about learning associations to complex stimuli. This contrast is whether an association to a complex stimulus is more effectively learned when only the whole stimulus is experienced or when its components are experienced separately. In the case of Chinese syllables, this complexity can be characterized as segment + tone components that are distributed in overlapping fashion across a brief time period of 300–400 milliseconds. Our aim was to develop and test specific hypotheses about how the components of syllable learning can be supported through decomposition of the syllable into its components.

Background

The Mandarin Chinese tone system has five tonal values: high-level (often labeled as 1), rising (2), low-falling-rising (3), high-falling (4), and mid-flat (neutral, 5). Tonal information is crucial for comprehending spoken Chinese. A change in the tone alters the meaning of the syllable. For example, the syllable /ma/ can have four different meanings according to its tone represented as following: /ma/1-妈 (mother), /ma/2-麻 (hemp), /ma/3-马 (horse), and /ma/4-骂 (scold). Thus, a syllable’s meaning is determined by two kinds of phonemes:

segments (the /m/ and the /a/) and tones. (The tones are thus suprasegmental.) These two components jointly determine the syllable's semantic value as a meaning-bearing morpheme.

Children in Mainland China and the U.S. college students in this study learn pinyin, an alphabetic system, to assist with the reading of Chinese characters. Because the letters of pinyin directly map to segments, it is possible that, for learners, the pinyin supports the representations of the segmental phonology. Xu (1998) showed that all four Mandarin tones have a consistent alignment to the syllable regardless of internal syllable structure. Because the syllable (specifically, the rhyme or vowel ending) is a carrier of the tone, the processing of the syllable simultaneously involves segments and tone. This means that the spoken syllable is temporally integrated and can be difficult to decompose by nontonal language speakers (Lee & Nusbaum, 1993).

Acoustically, the four different tones are distinguished by the F0 contour slopes, or four different pitch shapes. Amplitude also differs among the four tones: Tone 3 has the lowest amplitude and Tone 4 the highest (Lin, 1965). Finally, the duration also differs: Tone 3 has the longest duration and Tone 4 the shortest (Chuang & Hiki, 1972). Among these three acoustic properties (pitch, amplitude, and duration), pitch is considered the defining feature, the critical factor that differentiates among the tones (Howie, 1976).

The challenge for second language (L2) learners is complicated by their relative lack of attention to the direction of pitch change, compared with native Chinese speakers (Gandour, 1983). Moreover, English listeners continue to fall behind Chinese listeners after five sessions of training in a discrimination task of Thai tones (Wayland & Guion, 2004). Even for American college students who have been learning Chinese language for one semester, Wang et al. (2003) found that they encountered great difficulty in acquiring tone skill. Wang et al. used a matching task to test beginning Chinese learners' phonological processing skills. When matching was based on perceiving the same tone in two syllables, the Chinese learners showed poorer performance than when the matching was based on syllable onset and rhyme. This result can be explained by the fact that onset and rhyme are highly general phonological components that are shared between English and Chinese, whereas tone is a feature of Chinese but not English. Our hypothesis is that the difficulty arises from the demands of attending to complex stimuli (syllables with tonal and segmental components) in which at least one of the components is unfamiliar (i.e., not represented in the learners' native language). Adding to the difficulty in this case is that the set of Chinese segmental phonemes include some that are not present in English. A stable perception of the tone—one that transfers to other syllables—is at risk while the hearer tries to extract the segments from the syllable.

There have been few studies directed at training listeners to perceive tones. In one study, Wang, Spence, Jongman, and Sereno (1999) used an auditory training procedure to train U.S. listeners to perceive Chinese tones. The four Chinese tones were trained in all possible paired combinations. In each pair, the same segment with different tones was presented to the learner sequentially. The training procedure followed the order from easy pairs (Tones 1 and 3 or Tones 1 and 2) to more difficult pairs (Tones 2 and 3 or Tones 1 and 4). The posttest results showed a significant increase of identification accuracy from the pretest. These results show that training through exclusively auditory input can be effective. It is possible, however, that multimodal training that combines visual and auditory training can be even more effective, especially if the use of two modalities is designed to support learner attention to tone information.

The value of well-designed multimodal displays is seen in a variety of learning contexts, especially higher level conceptual learning (Mayer & Moreno, 2002). Although the tone learning problem is primarily perceptual rather than conceptual, some of the conclusions from Mayer and Moreno (2002) seem relevant, especially the idea that simultaneous presentation of information across modalities is more effective than asynchronous presentation. In the case of tone learning, providing the segmental information in one modality and the tone information in the other modality simultaneously may allow modality-specific attention to tone. Certainly, attending to relevant features is important in learning speech categories. For example, by directing the target of learning to either the sound or the meaning of Hindi words, Guion and Pederson (2007) found that the sound attending group did better than the semantic attending group in phonetic learning, suggesting attention to sound features facilitates the acquisition of novel phonetic categories. We suggest that the principle that operates here is one of attentional focus and add a corollary: When the input within a single modality (e.g., auditory) has complex overlapping components, separating these components across modalities allows a modality-specific attention (i.e., visual attention, auditory attention) to be directed to the relevant features of one of the components (e.g., tone). In more concrete terms, presenting the tone in a visual form separately from segments directs the listener's visual attention to pitch information and auditory attention to segment information.

The alignment of visual information with tone, in fact, has been well established in Chinese. Mandarin tonal pitch was found to be highly correlated with the shape of the F0 contour in early linguistic studies (Bai, 1934; Chao, 1930, 1968). Representing pitch information in the visual modality can benefit from the congruence effect in cognitive task performance in the visual modality (Glenberg & Kaschak, 2002). The natural congruence between an auditory pitch contour and its visual analog is that higher pitch is represented upward and lower pitch is represented downward. Rusconi, Kwan, Giordano, Umilta, and Butterworth (2006) demonstrated this kind of spatial-auditory congruence in participants' preferential pairings of keyboard responses to pitch. Participants responded significantly faster when high-frequency pitches were paired with an upper key (the number 6) and low-frequency pitches were paired with a lower key (the spacebar) than vice versa. This was true for both trained musicians and musically naïve participants.

As for the representation of segments, research on phonemic awareness suggests that pinyin helps developing the skill of segmenting Chinese syllables (Read, Zhang, Nie, & Ding, 1986). More generally, our hypothesis is that a visual representation of the spoken syllable can allow a discrete representation of the segments; that is, each letter corresponds to a single speech segment. This allows the learner to achieve a more stable representation of the spoken syllable. Thus, we hypothesize that providing explicit visual information in the form of pitch contours (for tone) and pinyin (for segments) in coordination with a spoken syllable will facilitate learning.

In the present study, we tested a novel training approach designed to facilitate students' perception of Chinese tones. More specifically, two basic principles of instruction are tested:

1. *Componential*: Because a tonal syllable that also contains unfamiliar segments is perceptually complex, learning is supported by presenting the components (syllable and tone) separately.
2. *Dual modality*: One source of information concerning the component comes from spoken input and one comes from visual input.

The componential aspect of the training is to draw learners' attention to the critical features of a syllabic tone by providing pinyin. The dual-modality feature allows the learner to attend specifically to a visual representation of tone and extract critical tonal contour information and then to add this information to the representation of syllable phonology.

To test training regimes that implement this idea to different degrees, we designed three training conditions in which we trained students to perceive tones using (condition 1) visual pitch contours that depict the acoustic information of the tones, together with pinyin spelling of the spoken syllable; (condition rhyme 2) numbers that represent the tones, together with pinyin spelling of the spoken syllable; and (condition 3) visual pitch contours, without pinyin spelling.

Condition 1 implements the instructional principle fully; condition 2 lacks the dual modality feature; and condition 3 lacks the redundancy provided by the pinyin spelling. Therefore, condition 1 (i.e., Contour + Pinyin) can be considered our experimental condition and the other two (i.e., Number + Pinyin, Contour Only) can be regarded as control conditions.

A shared feature of all three conditions is the auditory spoken syllable, which provides a combination of segmental and tonal information. The pinyin spelling grounds the segmental perception of the syllable in graphic form. It should support the learner in representing the spoken phonemes of the syllable. The pitch contour grounds the tonal perception of the syllable in a visual form. It should aid the learner in representing the tone heard in the syllable.

Because of limitations in the number of students available for the study, we could not carry out the ideal (2 × 2) design that fully crosses pinyin (present/absent) and pitch (contour vs. number representation). Instead, we test two hypotheses by making two comparisons among the three conditions: (a) condition 1 versus condition 2 tests whether visual tonal pitch contour facilitates students' perception of tones conditioning on pinyin presented; (b) condition 1 versus condition 3 tests whether access to pinyin spelling facilitates the perception of tones when pitch contour is presented. Furthermore, intact classes instead of individual students were used as the randomization unit. We matched the classes as much as possible. More details are provided in the Participants section.

Note that visual cues of tone can be presented above a nonmedial vowel as a diacritic mark that illustrates the tone's general contour (flat, rising, fall-rising, and falling). (A nonvisual manner of marking tones with numbers is also common in computer-assisted learning.) However, the goal of the present study was not to test whether a real pitch contour or an illustrative diacritic serves better but to test a general hypothesis that related visual information can facilitate perceptual learning.

The present study is part of a larger foreign language learning project associated with the Pittsburgh Science of Learning Center. An in vivo experimentation paradigm was used with curriculum vocabulary as training materials. We tracked students' learning in eight lessons across the first semester. In vivo experimentation allows the study of learning that occurs in actual classrooms during an actual course while also using laboratory-quality methods. In vivo experimentation has been used in disciplines such as algebra and chemistry and has provided very reliable results (Koedinger & Anderson, 1998; Yaron et al., 2001). Our use of this research paradigm aimed to test learning hypotheses based on the assumption that attending to the critical features of the pitch contour of the Chinese tone facilitates tone recognition.

Method

Participants

The participants in this study were 35 students enrolled in their first semester of an introductory Chinese language course at a U.S. university during the 2005–2006 school year. Most had English as their first language (L1), and two had Korean as their L1. Some of them had studied an L2 such as Spanish in high school. None of them had visual or hearing disabilities. Because the course included high levels of within-class student-student interaction, we concluded that a fully randomized assignment of students to conditions, in which students within the same classroom would be in different conditions, would risk contamination effects. Instead, each of the three classes at the same level was randomly assigned to one of the three training conditions. Because all three classes followed the same curriculum and were supervised by the same course director, any confound associated with class instruction was presumed to be minimal.

There were 10 students in the Contour + Pinyin condition, 16 students in the Number + Pinyin condition, and 9 students in the Contour Only condition. The students participated in the study as part of their online homework. The curriculum was organized to present one lesson each week. Therefore, the students went through the tone training online homework once a week after the classroom instruction.

Materials

Training materials for the online tutor were from the students' textbook. Eight lessons were involved in the training period. A total of 228 syllables, within 80 single-syllable and 74 bisyllable words, were used in the tutor. The number of syllables in each lesson ranged from 12 to 45. There was no syllable repetition. Supplementary words in the textbook were not included in the tutor. Students in the three training conditions were exposed to the same syllables in each lesson. The experimental manipulation was in the online tutor interface. Students in condition 1 (Contour + Pinyin) received visual pitch contours that depict the acoustic information of the tones, together with pinyin spelling of the spoken syllable. Students in condition 2 (Number + Pinyin) received numerical numbers that represent the tones as used in many online interfaces, together with pinyin spelling of the spoken syllable. Students in condition 3 (Contour Only) received visual wave forms but without pinyin spelling.

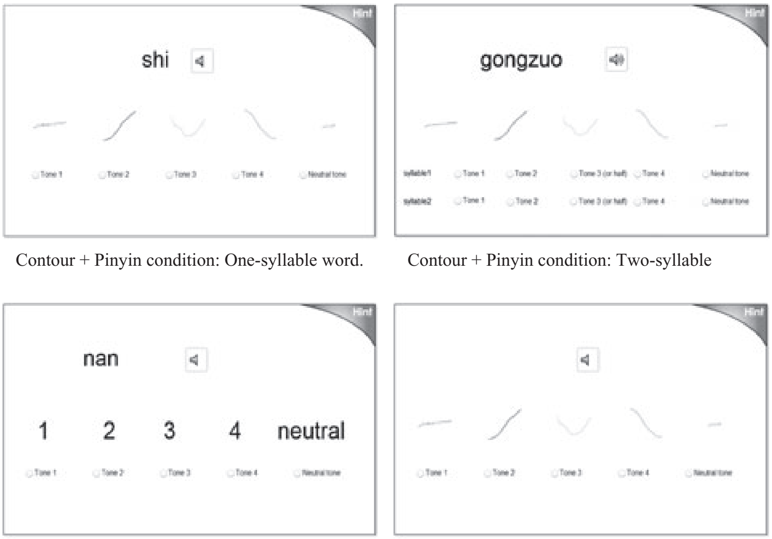
Online Tutor

The pitch contours were generated by slightly modifying the pitch contours of the syllable /ma/, produced by a female native Chinese speaker in the study of Luo, Gordon, Boemio, and Poeppel (2003). These contours are representative and match the pitch contours of different syllables, as contours do not vary much across syllables.

The three online tutors were implemented using Cognitive Tutoring Authoring Tools (Koedinger, Aleven, Heffernan, McLaren, & Hockenberry, 2004). The audio stimuli were digitally recorded by a Chinese language instructor at the same university from which the participants were recruited. When using the tutor, students went through the words one by one. First, they listened to the recorded audio of a monosyllabic or bisyllabic word, which was the same across the three conditions. Simultaneously, they viewed on the computer screen a different interface according to the condition to which they had been assigned. All students were asked to click on one of the five answer buttons representing the five possible tones: Tone1, Tone2, Tone3, Tone4, and Neutral Tone (the training materials contained syllables with neutral tones). In the case of bisyllable words, two sets of buttons were provided for the two syllables, as shown in Figure 1. The third-tone answer-button for a second syllable included the label "(or half)" in order to account for tone sandhi alternations when two third-tone patterns are combined in one bisyllable word.

The Hint System and Feedback

When an incorrect choice was made, a text box appeared to indicate that an error was made and directed the student to click on a "hint" button for helpful information. Students could click on (or ignore) each hint level in order. As shown in Table 1, the first-level help message encouraged the student to listen to the recorded syllable again and try to make another selection of the answer, whereas the second-level one specifically described the particular pitch contour of the given syllable. The third-level hint provided the correct answer and described the tone again. Students received another type of error message if the steps were attempted out of order, such as answering the second syllable of a two-syllable word before judging the first or attempting to proceed to the next word without correctly answering the current word. Positive feedback was given once the correct tone was selected.


Hint
shi
○ Tone 1
○ Tone 2
○ Tone 3
○ Tone 4
○ Neutral tone
gongzuo
syllable1 ○ Tone 1 ○ Tone 2 ○ Tone 3 (or half) ○ Tone 4 ○ Neutral tone
syllable2 ○ Tone 1 ○ Tone 2 ○ Tone 3 (or half) ○ Tone 4 ○ Neutral tone
nan
1
2
3
4
neutral

Contour + Pinyin condition: One-syllable word. Contour + Pinyin condition: Two-syllable

Number + Pinyin condition: One-syllable Contour Only condition: One-syllable word.

Figure 1 Examples of the computerized learning interface.

Pretests and Posttests

Two identical tone judgment tasks were carried out as part of the tutor that asked students to listen to a pair of syllables and judge whether the tones of the two syllables were same or different. The students in the three training conditions all completed this judgment task once before lesson 1 started as the pretest and once again after lesson 8 ended as the posttest. We decided to use this tone judgment task as the pretests and posttests (in addition to the tone identification task during the training sessions) to reduce the repetition. Both tasks, however, entail tone perception.

All of the syllables used in the two tests were real syllables in spoken Chinese and were learned in the first eight lessons; that is, the pretest and posttest materials were a subset of those used during the training. The syllable pairs consisted of three types: (a) The two syllables shared segmental information (both the onset and rhyme; e.g., shi1-shi3 with different tones); (b) the two syllables shared rhyme only (e.g., dao3-kao3 with same tones; and (c) the two syllables differed in both onset and rhyme (e.g., duo4-gong3 with different tones). The purpose for this manipulation was to see the difficulty levels associated with the segmental information. Twenty-four, 16, and 16 items were involved in each type, respectively, with 50% "yes" and 50% "no" responses.

Table 1 The hint system of the online tutor

Tone	Message No.	Hint message
All tones	(1st)	Focus on the pitch change and listen to the sound again.
Tone1	(2nd)	The sound pitch is nearly high-flat. Listen again.
	(3rd)	This is the first tone. Its pitch stays stable. Please click the correct answer.
Tone2	(2nd)	The sound pitch is high-rising. Listen again.
	(3rd)	This is the second tone. Its pitch goes up. Please click the correct answer.
Tone3	(2nd)	The sound pitch goes down at the beginning, then up at the middle (i.e., it is falling-rising). Listen again.
	(3rd)	This is the third tone. Its pitch goes down then up. Please click the correct answer.
Tone4	(2nd)	The sound pitch is going down (i.e., it is high-falling). Listen again.
	(3rd)	This is the fourth tone. Its pitch goes down. Please click the correct answer.
Neutral	(2nd)	The sound is short and its pitch is low-flat. Listen again.
	(3rd)	This is the neutral tone. It is short with little pitch change. Please click the correct answer.

There were some syllable repetitions in this test. Among the same onset and rhyme type of pairs, four syllables were repeated once (i.e., the participants heard them twice), eight syllables were repeated twice, and four syllables were repeated three times. Among the different onset and same rhyme type of pairs, two syllables were repeated once. There was no repetition among the different onset and different rhyme type of pairs. One talker recorded all of the stimuli. In the syllable pairs that shared same onset and rhyme and required a "yes" answer (e.g., shi1-shi1), the same token was presented in successive order.

The tests were embedded in the tutor. In the tests, the students heard a pair of stimuli, then selected one of the two buttons on the screen to indicate whether their tones were the same. Note that different talkers recorded the materials for the training and the prepost test materials.

Online Data Management

All activities done by the students were sent to the Pittsburgh Science of Learning Center data logging server. In the pretests and posttests, responses were logged and the accuracies were evaluated offline. During the online training, incorrect responses were recorded for each syllable.

Results

The data analyses consisted of two parts. In the first part, we analyzed the training log files to calculate the mean number of errors for each syllable in each lesson. The means, corrected for incomplete lessons and extraneous button clicks, were analyzed separately for each learning condition, producing three learning curves to be compared.

In the second part of the analyses, we examined the effect of training condition by comparing pretests and posttests. These comparisons test whether the experimental condition (i.e., Contour + Pinyin) improved student's tone perception more than the two control conditions (i.e., Number + Pinyin, Contour Only).

The Learning Curves

The average number of errors on all syllables in each lesson is plotted in Figure 2A. All three learning conditions showed an overall decrease of errors over the time, even though new syllables were introduced in each lesson. An exception to this trend was the increase in errors from lesson 3 to lesson 4 across all three conditions, although the materials of lesson 4 did not appear to be exceptional in any obvious way.

For a given training trial, the number of errors (maximum = 4) made on a syllable was not symmetrically distributed around the mean. Instead, the number of errors followed a near Poisson distribution, showing a higher probability of having zero and one error than two, three, or four errors. Thus, a linear regression, which has a normality assumption, is not appropriate for modeling the learning curves. Instead, a loglinear regression model was applied with the dependent variable being the number of errors for each syllable (Agresti, 2002). For example, the syllable "shi4" was tested on 9 subjects in the Contour + Pinyin condition, with seven errors made in total. The loglinear model took seven as the number of events (errors) and nine as the number of trials. After fitting the given dataset, the model solution provided an estimate of the number of errors for one trial under each treatment condition.

The independent variables included in the model were training condition as a treatment variable, lesson number (1–8) as a continuous variable, and Condition × Lesson interaction. The comparisons of Contour + Pinyin versus Number + Pinyin, and Contour + Pinyin versus Contour Only were planned comparisons, representing the main tests of learning conditions in our design. Number + Pinyin versus Contour Only was not directly compared because the confounding of two sources of experimental manipulation: number versus contour, and pinyin versus no pinyin.

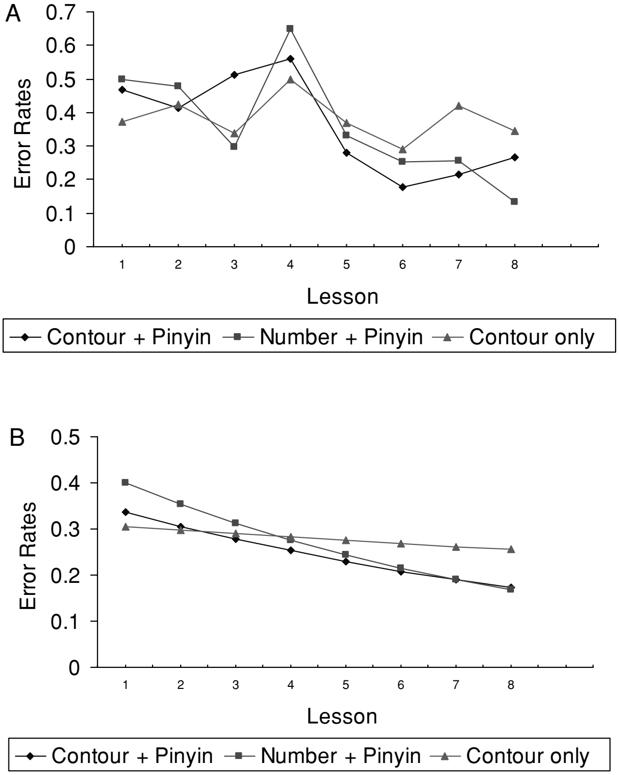


Figure 2 (A) Learning curves generated from the log data during learning; (B) Learning curves fitted by loglinear model.

The model can be written in the following form:

$$Log(error/student) = beta0 + beta1^{*}lesson + beta2^{*}C1 + beta3^{*}C2 + beta4^{*}lesson^{*}C1 + beta5^{*}lesson^{*}C2$$

C1 and C2 were dummy variables. C1 was assigned the value 1 for the Contour + Pinyin condition, C2 took the value 1 for Number + Pinyin condition, and both were 0 for the Contour Only condition. Thus, beta1 corresponds to the slope of the Contour Only condition in which both C1 and C2 were 0, beta1 plus *beta4* corresponds to the slope of the Contour + Pinyin condition, and *beta1* plus *beta5* is the slope of the Number + Pinyin condition. The six parameters *beta0* to *beta5* were estimated by statistical procedure GENMOD in the SAS software[1] (*beta0* = –1.166, *beta1* = –0.025, *beta2* = 0.174, *beta3* = 0.373, *beta4* = –0.071, *beta5* = –0.099). Because the parameters of a logarithmic equation do not have a transparent interpretation, Table 2 instead shows the

Table 2 Model estimates of the learning curves

	Mean errors	Slope of learning over lesson
Contour + Pinyin	0.236	−0.096
Number + Pinyin	0.253	−0.124
Contour Only	0.277	−0.025

estimated overall mean errors and the slopes based on the model. The estimated learning curves are shown in Figure 2B.

The mean errors showed a pattern of Contour Only > Number + Pinyin > Contour + Pinyin. The predetermined statistical tests on error rates based on the loglinear model showed a significant difference between Contour + Pinyin and Contour Only, $\chi^2(1) = 4.22, p = .04$, but not between Contour + Pinyin and Number + Pinyin, $\chi^2(1) = 0.85, p = .357$. The tests suggested that the participants made significantly more errors across the lessons under the Contour Only condition compared to the Contour + Pinyin condition.

However, because students were not randomly assigned to training conditions, the direct comparison of condition effects may seem questionable. Therefore, the learning curve slopes, based on the model parameters for the independent variable of lesson, were further compared to provide information on learning: A more negative slope indicates faster reduction of errors, thus better learning. All three slopes were negative, reflecting the overall reduction of errors over lessons. Although the slopes are caused by both classroom learning and online training, comparisons of slopes are good indicators of training differences because the classroom learning was identical across the three groups. Absolute slope values were ordered as Number + Pinyin > Contour + Pinyin > Contour Only. There was a significant difference between the slopes of Contour + Pinyin and Contour Only, $\chi^2(1) = 12.22, p < .001$. The difference in slopes between Contour + Pinyin and Number + Pinyin was not significant, $\chi^2(1) = 0.66, p = .415$. Number + Pinyin and Contour Only conditions were not compared directly. These results indicated that in the online learning activities presenting visual contour plus pinyin produced more rapid learning than presenting the contour only but did not produce more rapid learning than number plus pinyin.

The training materials contained both one-syllable and two-syllable words. The syllable position in the two-syllable words may have interacted with the training method on tone identification. In order to address the potential interaction among the lesson, training method, and syllable position, we categorized each training item into one of the three types of syllable position: syllable in a single-syllable word and first or second syllable in a two-syllable word. The effects of the syllable position and the interaction between the syllable position and training method as well as the other two-way and three-way interactions were added to the aforementioned loglinear regression model. Results showed that the single-syllable word had resulted in significantly fewer estimated errors per syllable (0.188) than the two-syllable word (error rates were 0.326 for the first syllable and 0.254 for the second syllable, respectively). Pairwise comparisons among the three types of syllable position were all significant (all $p < .001$). None of the two-way or three-way interactions was significant, however. The more accurate performance on the one-syllable word than the two-syllable word and the more accurate performance on the second syllable than the first syllable in the two-syllable word were probably due to the effect of working memory. The two-syllable words demanded more working memory than the one-syllable words and, therefore, were more error-prone. Furthermore, the second syllable in the two-syllable words was stored more recently in working memory than the first syllable and, therefore, could be identified more accurately.

We also compared the students' log time across the lessons under the three learning conditions. It is possible that the amount of time spent on online training varied across the conditions. The more time one spends in learning a language, the better one will become. Means (and standard deviations) of the log time in seconds were 198.14 (58.94), 203.81 (63.59), and 287.01 (148.87) respectively for the Contour + Pinyin, Number + Pinyin and Contour Only conditions. Pairwise comparisons with Tukey adjustment showed that the mean time of the Contour + Pinyin and Number + Pinyin conditions was close to each other ($p = .987$). Students in these two conditions spent less time than those in the Contour Only group; however, the differences were not statistically different ($p = .11$ and $p = .09$, respectively).

Pretest and Posttest Performance on Tone Judgment Tasks

The accuracies of tone judgments by the learning conditions and stimuli types are summarized in Figure 3. There were 8 subjects in Contour + Pinyin and 12 subjects in Number + Pinyin conditions who provided both pretest and posttest data. Due to the limited number of participants in the Contour Only condition who completed both pretesting and posttesting (only four), we did not include this third condition in these analyses. Both subject and item analyses were conducted. The item analysis has more statistical power for two reasons. First, the number of items was larger than the number of subjects. Second,

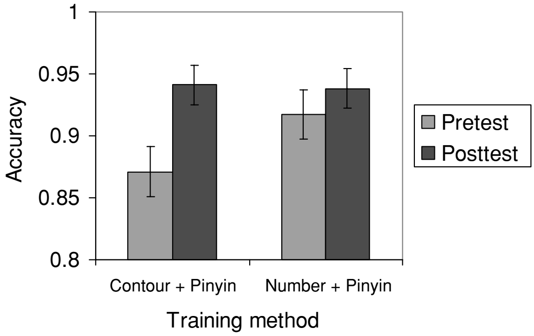


Figure 3 Accuracies of tone judgment task in the pretests and posttests. The error bars represent standard error of the mean; the *y*-axis ranges from 0.8 to 1.

the items were the same for Contour + Pinyin and Number + Pinyin conditions, which led to a repeated-measures analysis rather than a between-subjects analysis. Therefore, the item analysis is reported first.

Analysis of variance (ANOVA) was performed on the mean item accuracies with two repeated-measures factors: Test (pretest vs. posttest) and Condition (Contour + Pinyin vs. Number + Pinyin). Results showed a significant test effect, $F(1, 55) = 11.056, p = .002$, condition effect, $F(1, 55) = 6.892$, $p = .011$, and Test × Condition interaction, $F(1, 55) = 4.673, p = .035$. Scores were significantly higher in the posttest than the pretest: Accuracy increased from 87.1% to 94.1% for Contour + Pinyin and from 91.7% to 93.8% for Number + Pinyin. The improvement was significantly larger for Contour + Pinyin than Number + Pinyin (7% vs. 2.1%).

The ANOVA on mean subject accuracies includes pretest versus posttest as a within-subjects factor and Condition as a between-subjects factor. The results showed a significant test effect, $F(1, 18) = 7.552, p = .013$. However, the Test × Condition interaction, $F(1, 18) = 1.680, p = .211$, and the condition effect were not significant, $(F(1,18) = .982, p = .335)$.

Figure 3 showed that the pretest accuracy of the Contour + Pinyin condition was lower than the Number + Pinyin condition and that the posttest accuracies of both conditions nearly reach 95%. This raises the possibility that the greater improvement of the Contour + Pinyin condition could be a result of ceiling effect of the posttest and lower pretest accuracy in this condition. In order to statistically control for the effect of pretest performance, an analysis of covariance (ANCOVA) was performed, with the pretest-posttest accuracy improvement as the dependent variable, the experimental condition

Table 3 Means (and standard errors) of percent accuracy on tone judgment task by learning conditions and stimuli types for negative responses

	Contour + Pinyin		Number + Pinyin	
	Pretest	Posttest	Pretest	Posttest
Same segment (12 items; example: /shi/1-/shi/3)	96.3 (2.1)	98.2 (1.2)	98.6 (0.9)	100 (0.0)
Different onset, same rhyme (8 items; example: /dao/3-/kao/4)	73.6 (5.5)	87.5 (5.7)	82.3 (7.3)	83.3 (5.5)
Different onset, different rhyme (8 items; example: /duo/4-/gong/3)	91.7 (2.8)	93.1 (2.0)	92.7 (3.3)	92.7 (1.9)

as the repeated-measures variable, and pretest accuracy as the covariate. The key result was that even after controlling for the pretest scores, the effect of condition was still significant, $F(1, 54) = 17.317, p < .01$. The pretest-posttest improvement was greater for the Contour + Pinyin condition compared to the Number + Pinyin condition. The mean improvements adjusted for the covariate were 6.94% for Contour + Pinyin and 2.08% for Number + Pinyin. The covariate effect was significant, $F(1, 54) = 30.646$, $p < .01$. There was also a significant interaction between condition and covariate, $F(1, 54) = 14.651$, $p < .01$. The slope for the covariate was significantly different across the two conditions (–0.71 vs. –0.158), with the lower pretest score resulting in a higher pretest-posttest improvement.

Analysis of Segment and Tone Features of the Syllables

In the pretests and posttests, the syllable pairs had three different relations and their tones were either the same or different. The mean error rates by stimulus and response type are shown in Tables 3 and 4. To further explore the effect of these different items, while keeping the two repeated-measures factors (test and condition) in previous item-based ANOVA, two between-groups factors were added to the ANOVA: stimulus types (same segment, different onset-same rhyme, and different onset-different rhyme) and response type (yes vs. no). The results showed significant effects for Test, $F(1, 50) = 12.803$, $p = 0.001$, Condition, $F(1, 50)$, $p = .014$, and their interaction, $F(1, 50) = 5.948$, $p = .018$. Furthermore, the type effect, $F(2, 50) = 19.925$, $p = .000$, and the Response × Type interaction, $F(2, 50) = 8.758$, $p = .001$, were both significant. The response type effect was not significant ($F < 1$), nor was there a three-way interaction.

Table 4 Means (and standard errors) of percent accuracy on tone judgment task by learning conditions and stimuli types for positive responses

	Contour + Pinyin		Number + Pinyin	
	Pretest	Posttest	Pretest	Posttest
Same segment (12 items; example: /shi/1-/shi/1)	95.4 (1.7)	99.1 (0.9)	97.9 (1.1)	100 (0.0)
Different onset, same rhyme (8 items; example: /dao/3-/kao/3)	86.1 (5.0)	93.1 (3.6)	89.6 (6.0)	94.8 (2.2)
Different onset, different rhyme (8 items; example: /duo/4-/gong/4)	70.8 (4.2)	88.9 (3.6)	82.3 (5.1)	85.4 (3.4)

The test effect revealed that the participants had higher accuracies on the posttest than pretest across different types of materials. Among the three types of stimuli, the students had the lowest accuracies on the item pairs that had the same rhymes but demanded negative responses (/dao/3-/kao/4) and on those pairs having different segments but demanding positive responses (/duo/4-/gong/4). The students had the highest accuracy rates on the items involving same segment regardless of whether the tone required a same or different response.

Because there was no significant three-way interaction, it is hard to attribute the two-way interaction of test and condition to any specific type of materials. However, Tables 3 and 4 suggest that the main source of improvements for Contour + Pinyin came from the two most difficult types of pairs (same rhyme but different tone and different segments, but same tone): from 73.6% to 87.5% and from 70.9% to 88.9%, respectively.

Use of the Hint System

The students' use of the hint system was also logged. However, hints were not used very often. The average numbers of hints used in the three conditions were 0.02, 0.01, and 0.04 per syllable by student, respectively. A repeated-measures ANOVA on the use of hints with condition as the factor showed no significant difference among the three conditions, $F(2, 14) = 1.005, p = .362$; Greenhouse-Geisser $\varepsilon = 0.604$.

Discussion and Conclusion

This in vivo experiment provided log files that reflect the online learning process. By analyzing the log files and the two tests carried out online, we obtained evidence for the effect of utilizing visual information in identifying tone in Chinese. The results showed that beginning Chinese learners experienced better learning (fewer errors) over the course of detecting tones of syllables in eight different lessons of their Chinese curriculum when trained with visual displays of pitch contour information together with pinyin spellings than when supported with a visual contour only. With pitch contours plus pinyin, learners also showed better improvement in their tone judgment performance from pretest to posttest when compared to numerical signs plus pinyin. These results suggest that pinyin spellings lead to a more rapid improvement in tone recognition during the instructional period, whereas visual contours lead to more improved performance on the posttest of tone perception after learning. We discuss these two conclusions in turn.

First, the effect of pinyin spellings was to increase the rate of tone learning, whether the pinyin was combined with visual contours or numerical indicators. In fact, the performance of these two conditions was close to each other during the course of the semester. This general effect of pinyin can be explained by our attention assumption: That presenting the components of a complex perceptual event separately allows attention to focus on one of them. The spoken Chinese syllable is an integral stimulus, with vowel and tone information overlapping. The task we gave to the participants required them to select a pitch contour out of the integrated features of the syllable. In both training conditions using pinyin, pinyin was presented visually before the corresponding syllable was heard. Thus, pinyin served to help fix a representation of the segmental phonemes in mind and further allowing attention to focus on the tone as the spoken syllable was presented.

The facilitation due to a pitch contour relative to numeric representation of tone may reflect a modality principle of the sort observed in higher level learning (Mayer & Moreno, 2002). Specifically, the better training effect from the Contour + Pinyin condition in comparison to the Number + Pinyin condition is most likely due to fact that the Contour + Pinyin condition provided both visual and auditory information of the pitch contour of the tone, hence resulting in a more robust representation that was later easily recalled. Again, because tone learning is essentially a perceptual learning task, more specific interpretations of dual-modality effects must also appeal to more basic processes such as perception and attention. In those terms, the visual contour information was available to draw attention to the auditory pitch information of the tone, supporting a percept of complex tonal information for learners of Chinese as a second or foreign language. The congruence of rising and falling acoustic pitch with rising and falling graphic contours allows an iconic, nonresource demanding representation of pitch information, relative to the arbitrary numeric indicators. In effect, the visual pitch contour, which was emphasized in the display and the hint system, was able to draw the learners' attention to the F0 slopes, addressing the lack of attention to the direction of pitch change by English learners of Chinese (Gandour, 1983).

More generally, our results also are related to recent studies that suggest that pitch information can be represented in spatial terms by both trained musicians and musically naïve participants (Rusconi et al., 2006). Our findings provide new evidence that representing pitch in spatial terms can occur for novel and complex auditory stimuli. Similar to aligning upper keys with higher pitches leading to faster reaction times, as in Rusconi et al. (2006), visual pitch contours are isomorphic analogues to spoken pitch contours and the cognitive system can take advantage of a natural congruence between spectral and spatial processing of auditory stimuli.

It is important to recognize that syllable perception in Chinese presents a complex stimulus and that discrimination among temporally integrated features (segments and tones) is a problem. Evidence for this comes from our analysis of test pairs in the tone judgment task, whose results are shown in Tables 3 and 4. By analyzing the posttest tone judgment data, we found that the participants made more errors when rejecting the /dao/3-/kao/4 than rejecting the /shi/1-/shi/3 pair (Table 3). This result that tone discrimination is easier when both onset and rhyme are the same suggests that the beginning learners had difficulty in separating the tone from the syllable segments. It is possible that the learners were distracted by the different onsets, which made it harder to perceive that the tones were also different between /dao/3 and /kao/4. Due to their limited Chinese language proficiency, the learners' tonal perception was conflated with their segmental perception. Interference was also shown in the cases when the stimulus pairs did not share the rhymes. In the context of hearing different onsets and rhymes in /duo/4 and /gong/4 (Table 4), the learners found it difficult to notice their identical tones. Overall, the learners had difficulty in extracting the tonal information independent of the syllable segmental information. It is interesting that these two difficult types of syllable pairs received the most significant improvement in the tone judgment task in the visual contour plus pinyin training condition. We suggest that presenting both visual contour and pinyin emphasizes the separation of segmental and tonal information, thus facilitating the development of independent representation of the two types of information.

Our results may also add to evidence showing that adult learners can acquire, with training, difficult linguistic contrasts that are learned naturally by native speakers during childhood. McClelland and his colleagues have successfully taught Japanese adults the /r/-/l/ discrimination in English, a nonnative speech contrast, after several sessions of training (e.g., McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; McClelland, Fiez, & McCandliss, 2002). In our study, the issue was also a phonemic contrast (although, here, a suprasegmental contrast) not present in the native language. Acquiring this contrast is difficult for an L2 learner, even with training. We do not suppose that our training procedure is the optimal one, and we would expect to see even more robust training developed based on speech practice as well as componential input. However, as in the McClelland et al. (2002) /r/-/l/ training, a procedure that draws attention to the focal point of contrast—an onset contrast in their case and a tonal contrast in ours—is important. It appears that visual presentations of tone information are an effective means to do this in Chinese L2 instruction.

It is important to note that the tone judgment task used in the pretests and posttests is a type of tone discrimination task, whereas the tone training or instructional task is a type of tone identification task. Even though both involved tone perception, the mechanisms underlying the two tasks are different. Logan, Lively, and Pisoni (1991) suggested that phonetic training (e.g., vowel, consonants, or lexical tones) using the identification task, in comparison to the discrimination task, encourages learners to rely more on phonetic information stored in long-term memory rather than on rapidly fading perceptual information in short-term memory. Wayland and Li (2008) further stipulated that the identification training task draws learners' attention to "inclusionary" features (e.g., which category does the tone fall into?), whereas the discrimination training task draws attention to both "inclusionary" and "exclusionary" features (e.g., why should the two tokens fall into the same or different phonetic category?). It is possible that our tone identification training had a facilitative effect on tone discrimination in the posttest because both tone identification and tone discrimination involved the attention to the "inclusionary" features of the tone category. It is also possible that tone identification training may not be the most effective training regime in promoting tone discrimination, as the tone identification training may have only encouraged the attention to the "inclusionary" feature of the tone category.

The tone discrimination task materials used in the pretests and posttests were recorded by the same talker; in particular, the same token was used in the items that shared the onset and rhyme and required a "yes" response. This type of physically identical stimuli pairs may allow the participants to make their decision based on the fact whether the stimuli has been presented twice, rather than on their judgment whether the same phonetic category has been presented (see Wayland & Li, 2008, for a discussion). Future research could consider including categorical same/different discrimination materials in which the stimuli in each pair are always physically not identical (e.g., when the stimuli pair were recorded by two different talkers). The categorical materials tap into learners' ability to ignore acoustic variations and to draw more on perceptual constancy. Hence, it would be interesting to see the effect of training method on categorical same/different discrimination.

Finally, some limitations in our design are worth noting. First, the number of participants in each experimental group was small ($N = 10$, 16, and 9 for the Contour + Pinyin, Number + Pinyin, and Contour Only condition, respectively). Future research needs to increase the sample size in order to improve the statistical power. Second, although the three intact classes recruited for the three experimental conditions were matched as closely as possible, the different instructors may still exert some influence on students' learning. Furthermore, additional standardized measures could be administered prior to experiments in order to obtain more information regarding the comparability among the three classes in terms of their overall intellectual ability and language learning aptitude. If it is feasible, future research needs to incorporate a design using a within-class random assignment or rotating instructors at set times during the experiment to avoid the confound resulting from the different instructors or classes.

Revised version accepted 9 February 2010

Note

[1] The data analysis for this article was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

References

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley-Interscience.

Bai, D. (1934). Guanzhong shengdiao shiyan lu [Experiments with tones of Guanzhong dialects]. In *Guoli Zhongyangyanjiuyuan Lishiyuyan Yanjiusuo Jikan* [A collection by Academia Sinica Institute of History and Language](pp. 355–361). Nanjing, China: Academia Sinica.

Chao, Y. R. (1930). A system of tone letters. *Le Maitre Phonetique*, *30*, 24–27.

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

Chuang, C. K., & Hiki, S. (1972). Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *Journal of Acoustical Society of America*, *52*, 146.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, *9*, 558–565.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 57–77). Amsterdam: Benjamins.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.

Koedinger, K. R., Aleven, V., Heffernan. T., McLaren, B., & Hockenberry, M. (2004). Opening the door to non-programmers: Authoring intelligent tutor behavior by demonstration. In the *Proceedings of 7th Annual Intelligent Tutoring Systems Conference*. Berlin: Springer-Verlag.

Koedinger, K. R., & Anderson, J. R. (1998). Illustrating principled design: The early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments*, *5*, 161–180.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics, 53*, 157–165.

Lin, M. C. (1965). The pitch indicator and the pitch characteristics of tones in standard Chinese. *Acta Acoustics (China)*, *2*, 8–15.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustic Society of America*, *89*, 874–886.

Luo, H., Gordon, M., Boemio, A., & Poeppel, D. (2003, March). *The perception of FM sweeps by Chinese and English listeners*. Paper presented at the Annual Conference of the Cognitive Neuroscience Society, New York.

Mayer, R., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*, 107–119.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the /r/-/l/ contrast to Japanese adults: Test of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, *2*, 89–108.

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

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sound depends on know alphabetic writing. *Cognition*, *24*, 31–44.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*, 113–129.

Wang, M., Perfetti, C. A., & Liu, Y. (2003). Alphabetic readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies in Reading*, *72*, 183–207.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American liteners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3469–3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*, 681–712.

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

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179–203.

Yaron, D., Freeland, R., Lange, D., Karabinos, M., Milton, D. J., & Belford, R. (2001). *Uses of a flexible virtual laboratory simulation in introductory chemistry courses*. Paper presented at online conference CONFCHEM (CONFerences on CHEMistry): On-Line Teaching Methods, American Chemical Society. Retrieved April 18, 2007, from http://www.ched-ccce.org/confchem/