Learning Words with Lexical Tone: Is Manipulation of Attentional Focus Beneficial?

Wenyi Ling and Theres Grüter*

Word learning is one of the initial and essential parts in learning a language. As building blocks of language, words involve both form and meaning. In all languages, words are composed of consonants and vowels, while in some languages, tones also play a comparable role in building words (Yip, 2002). Previous studies show that speakers of non-tonal languages (e.g. English) find it difficult to learn words in a tonal language (e.g. Pelzl et al., 2019). A number of factors are likely to contribute to success in learning tones. The most widely recognized factors contributing to variation in learning outcomes among second language (L2) speakers include tone experience, pitch ability, musicality or experience with music, L2 aptitude and general cognitive ability. Among these factors, experience with tones in the native language (Chan & Leung, 2019) and learners’ ability to perceive pitch patterns in a non-verbal context showed the highest correlation with tone learning success (e.g. Bowles et al, 2016; Wong & Perrachione, 2007).

While manipulation of the factors listed above is mostly beyond learners' control, teachers and educators are seeking ways to improve learning outcomes by focusing on external factors that can be manipulated, such as different teaching and training methods. Applied linguistic studies have reported the effectiveness of using various teaching and training methods, such as visualization of tone contours (Liu et al., 2011), using color or number coding (Godfroid et al., 2017) and music (Lin, 1985), hand gestures or other body movements (Tsai, 2011) in the perceptual learning of tones. Common to all of these methods is that they try to draw learners’ attentional focus to tones as a minimally contrastive feature. This aligns with a number of different theoretical models of language learning. For instance, the Noticing hypothesis claims that “noticing is necessary for intake” (Schmidt, 1990, p. 141). If learners do not pay attention to contrastive features of a cue, they could not intake the information and learn the cue. According to the Automatic Selective Perception model (Strange, 2011), the perceptual salience of a cue is influenced by speakers’ linguistic experience, but experimental manipulation can potentially reallocate

*Wenyi Ling, University of Hawai’i at Mānoa, wenyi9@hawaii.edu; Theres Grüter, University of Hawai’i at Mānoa, theres@hawaii.edu. We are grateful to Patrick Wong and Tyler Perrachione for sharing the PCPT from Wong & Perrachione (2007) with us. This study was supported by a Doctoral Dissertation Research Improvement grant from the National Science Foundation (BCS-1824082).
their attentional focus. Finally, the Competition Model (MacWhinney, 2005, 2012) supports cue-focus training more directly by arguing that presenting the contrastive form can increase the relative strength of a cue in acquisition.

Focusing learners' attention on the contrastiveness of a cue such as tone thus seems not only intuitively helpful, but also appears broadly supported by existing theoretical models. It is important to note, however, that very few studies have directly tested this assumption. Nevertheless, many researchers in applied linguistics accept this assumption as a premise, and focus on the examination of different types of cue-focus training methods in the learning of tones. In a study by Liu et al. (2011), for example, L1-English first-year learners of Mandarin were trained to learn Mandarin tones on syllables in three different learning conditions – (a) contour + pinyin, (b) number + pinyin, (c) contour only – in an actual classroom setting. All three of these learning conditions are versions of cue-focus training, and the study is based on the assumption that cue-focus training could direct learners’ attention to the critical features of tones and thus lead to better learning. Learning outcomes were measured in terms of decrease in error rates on two identical tone judgment tasks conducted as pre- and post-tests. Results showed that the contour + pinyin condition had more error reduction than the other two conditions. This finding is of potential pedagogical relevance, but it does not constitute direct support for the assumption that cue-focus training is effective in principle in the learning of tones since all three conditions focused directly on the contrast of tones. Furthermore, additional confounds between groups were difficult to control since the study was conducted in an existing classroom setting.

Based on Liu et al.'s (2011) positive results with dual visual representation, Godfroid et al. (2017) compared the effectiveness of five multimodal methods (three single-cue methods: number, color and pitch contour; two dual-cue methods: color + number, color + pitch contour) of tone contrastive training for Mandarin tone perception in a more controlled experimental setting. Results from pre-test as well as both immediate and delayed post-tests showed that all training methods were effective, with more advantage with pitch contour and number than color, while dual-cue methods were no more effective than single cue-methods. Godfroid et al.'s (2017) findings align with Liu et al.'s (2011) in that training could improve learners’ perception of tones. Neither of those two studies, however, tested the assumption that cue-focus training is effective in learning of tones in comparison to some baseline condition, and both studies only examined the learning effect at a phonological level.

Though previous studies provide ample evidence that the accuracy of tone identification by listeners with non-tonal language backgrounds can be greatly improved even after short-term training (Wang et al., 1999; Wong & Perrachione, 2007; Liu et al., 2011; Wang, 2013; Godfroid et al., 2017), L2 speakers' actual and long-term difficulty with tone processing appears to lie at the lexical level (Qin, 2017; Pelzl et al., 2019). In Pelzl et al.’s (2019) study, the accuracy of tone identification by English-speaking learners of Mandarin was similar to that of native listeners, while learners were less successful than native
listeners at rejecting non-words in a lexical decision task when the non-words and words differed only by tone, indicating a “disconnect between L2 abilities to categorize tones as phonetic objects and abilities to utilize those categories as lexical cues” (Pelzl et al., 2019, p. 69). As a lexical cue, the most important and fundamental function of tones that learners need to acquire is how to use them along with segmental cues to recognize words. Previous research shows that English speaking L2 learners of Mandarin do not appear to have developed sufficiently automatized selective perception routines that allocate tone the weight it has in L1 Mandarin processing (Ling & Grüter, 2018), which might be due to the weaker cue strength of tones than segments for L2 learners. As suggested by the Competition Model (MacWhinney, 2005, 2012), presenting contrastive forms of the tonal cue might increase its relative strength in acquisition, which is the assumption of the effectiveness of cue-focus training. If this assumption is correct, listeners trained in a cue-focus condition should outperform those trained in a no cue-focus condition in learning words with tones. The goal of the present study is to test this prediction.

1. The current study

Realizing the necessity of examining the assumption that cue-focus training is beneficial, and recognizing the gap between vocabulary teaching practices and theoretical models of learning, the current study investigates the effectiveness of cue-focus training in word learning in a controlled laboratory setting. To this end, we drew on the methodology and design of previous laboratory-based studies on word learning in tonal languages. First, to avoid the influence of statistical regularities associated with the distribution of tones in natural languages (Wiener et al., 2019), we use novel words in an artificial language (Wong & Perrachione, 2007; Hayakawa et al., 2019). Second, a single-session laboratory-based auditory novel word-learning study design (Quam & Creel, 2017) enables us to test the immediate learning outcome after short-term training. Third, other potentially confounding factors at a participant level are controlled by only including English speakers without previous experience of any tonal language, randomly assigning them to one of three training conditions, and independently assessing their pitch ability prior to the experiment (Wong & Perrachione, 2007). Within this overall study design, any differences in learning outcomes between the three training groups are likely to be attributable to the experimental manipulation, i.e., training group. In sum, this study is designed to directly examine the assumption that manipulation of attentional focus by presenting contrastive cues during training benefits the learning of words with lexical tones.
2. Methods
2.1. Participants

Data from a total of 90 self-identified English native speakers (male = 26, other = 1; mean age: 22, range: 18-47) recruited from the University of Hawai‘i community and randomly assigned to one of the three training groups (see 2.3.1), was included for analysis. An additional 9 participants took part, but their data was excluded due to the participants knowing a tonal language (3: Cantonese, Thai and Vietnamese) or technical problems (6). All participants reported normal hearing and normal or corrected vision. Information on professional music experience and language experience was collected as part of a background questionnaire. All except 11 of the remaining 90 participants reported learning experience of another second language. None of those languages was a tonal language and none of the participants had professional music experience. The study protocol was approved by the Institutional Review Board at the University of Hawai‘i, and participants were compensated with extra course credit or a small amount of money.

2.2. Materials

2 consonants (/p/ and /s/), 3 vowels (/a/, /u/, /i/) and 3 tones (rising, level, falling) were used to create 18 novel words, simultaneously comprising 6 triplets minimally contrastive by vowel (e.g. /pa/-rising, /pu/-rising, and /pi/-rising) and 6 triplets minimally contrastive by tone (e.g. /pu/-rising, /pu/-level, and /pu/-falling). Table 1 shows all 18 words and their associated meanings.

Table 1. The artificial vocabulary

<table>
<thead>
<tr>
<th>/pa/-rising (flower)</th>
<th>/pu/-rising (house)</th>
<th>/pi/-rising (knife)</th>
<th>/sa/-rising (nose)</th>
<th>/su/-rising (pants)</th>
<th>/si/-rising (fork)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pa/-level (cup)</td>
<td>/pu/-level (shoe)</td>
<td>/pi/-level (tree)</td>
<td>/sa/-level (book)</td>
<td>/su/-level (hand)</td>
<td>/si/-level (plate)</td>
</tr>
<tr>
<td>/pa/-falling (fire)</td>
<td>/pu/-falling (bag)</td>
<td>/pi/-falling (hat)</td>
<td>/sa/-falling (ball)</td>
<td>/su/-falling (pen)</td>
<td>/si/-falling (melon)</td>
</tr>
</tbody>
</table>

Note: Each novel word, written in the International Phonetic Alphabet, is followed by its tone and the associated meaning in English in parentheses.

The choice of consonants, vowels and tones was based on distinctiveness and familiarity. The bilabial stop /p/ and the alveolar sibilant /s/ are different in both manner and place of articulation. Low front /a/, high front /i/ and high back /u/ are the three most distinctive vowels and are common across languages. All consonants and vowels are familiar to English speakers. The three tones are similar to pitch patterns of the Mandarin rising, level and falling tone, but synthesized with the same duration and intensity. The Mandarin dipping tone (or low falling rising tone) was not included because it is reported as the most
confusing tone for both L1 and L2 speakers of Mandarin (e.g. Hao, 2012; Pelzl et al., 2019). Since the purpose of this study is to test the effectiveness of cue-focus training, we tried to match the perceptual difficulty between segments and tones. Though English does not have lexical tones, English speakers are generally familiar with rising, falling and level pitch as intonation markers (e.g. So & Best, 2010).

The meaning of each word was chosen to be easily imageable and common. To ensure participants would associate acoustic word forms with concepts rather than specific visual images, a set of clip-art images and a set of photos depicting the same concepts were selected and used as visual stimuli for the training session and the test session, respectively (see Figures 1 and 2 below).

A female native speaker of Mandarin and a male L1-English L2 learner of Mandarin were asked to pronounce the 6 words with level tones in isolation slowly and clearly with normal volume in a sound-proof booth three times, and were recorded via a built-in microphone in a Mac Pro computer at 44.1kHz using Praat (Boersma & Weenink, 2016). One of the three tokens for each word was selected from each talker according to the sound quality. The selected soundfiles for each word and talker were normalized by intensity at 80 dB using Praat. Pitch patterns were interpolated linearly through each stimulus, using the PSOLA method implemented in Praat. Following the auditory manipulation in Wong and Perrachine (2007), the pitch contours in this study were also modeled on the values obtained by Shih (1988). Each pitch pattern changed linearly from the beginning point to the end point. For each minimal tone triplet from each talker, the average fundamental frequency (F0) of the level tone was used as the starting and ending points for level tone. Based on the word with level tone, the pitch contours of rising and falling tone in each triplet were generated according to the scheme in Table 2.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Beginning point</th>
<th>End point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>F0_L</td>
<td>F0_L</td>
</tr>
<tr>
<td>Rising</td>
<td>0.74*F0_L</td>
<td>F0_L</td>
</tr>
<tr>
<td>Falling</td>
<td>1.1* F0_L</td>
<td>0.385*F0_L</td>
</tr>
</tbody>
</table>

Thus, except for F0, all other acoustic parameters (including duration and voice quality characteristics) were consistent in each minimal tone triplet. Stimuli produced by the female speaker were used for the training session and stimuli produced by the male speaker were used for the test session to ensure listeners would associate the phonological representation of each word with its concept instead of relying on specific acoustic details.

2.3. Procedure

After completing a web-based questionnaire about basic demographic information and language experience, all qualified participants came to the lab...
for a single experimental session that lasted approximately one hour. First, participants completed a pitch contour perception test (PCPT; Wong & Perrachione, 2007; Perrachione et al., 2011) presented in PsychoPy2 (Peirce, 2007). This task consists of 120 trials (4 talkers * 5 vowels * 3 pitch contours * 2 repetitions) in which a sound file is played while two arrows depicting possible pitch contours (level↑, rising↗, or falling↘) are presented on the screen. Participants were instructed to press the corresponding button on the keyboard (“1” for the left arrow, “2” for the right arrow). No feedback was provided. The task was self-paced (see Wong & Perrachione, 2007; Perrachione et al., 2011 for further details). After completing this task, the word learning experiment started.

2.3.1. Training session

Participants were randomly assigned to one of three training groups: Tone-focus, Vowel-focus and Control group. All participants were told that they were going to see images and hear them named; they needed to repeat the words, try to learn them and would be tested later. In the Tone-focus group, three words with the same segments but different tones (e.g. /pu/-rising, /pu/-level and /pu/-falling) were presented on one slide (Figure 1a). In the Vowel-focus group, three words with the same consonants and tones, but different vowels (e.g. /pa/-rising, /pi/-rising and /pu/-rising) were presented on one slide (Figure 1b). In the Control group, three words that differed in more than one cue (e.g. /pu/-rising, /sa/-falling and /si/-level) were presented on one slide (Figure 1c). Each time a new slide appeared on the screen, an arrow would point at the first image while it was being named. Participants were instructed to press the space bar to move the arrow to the second and third images so that they would be named in turn. Each slide was presented six times in six rounds, and within each round, the order of presentation was randomized, for a total of 36 training trials.

![Figures](a) Tone-focus group  (b) Vowel-focus group  (c) Control group

**Figure 1. Examples of triads in 3 training groups**

2.3.2. Test session

Following the training session, all participants were tested in the same 2-alternative forced choice task. On each trial, participants were presented with images of two objects while hearing one of them named and were asked to click on the named object. For each word (e.g. /pu/-rising), there was one trial with a competitor differing from the target only by tone (Tone-pair trial: /pu/-falling, Figure 2a), one with competitor differing from the target only by vowel (Vowel-pair trial: /pa/-rising, Figure 2b), one with a competitor differing from the target
only by consonant (Consonant-pair trial: /su/-rising, Figure 2c) and two trials with words sharing no phonological similarities with the target (Baseline: e.g. /si/-level, Figure 2d). Overall, the test session consisted of a total of 90 trials (18 words *5) with each object appearing as target and competitor five times.

(a) Tone-pair trial
(b) Vowel-pair trial
(c) Consonant-pair trial
(d) Baseline trial

Figure 2. Examples of different trial types

3. Results

Data from 31 participants in the Tone-focus group, 31 in the Vowel-focus group and 28 in the Control group were included in the final analysis.

3.1. Pitch contour perception test (PCPT)

As discussed above, the PCPT was included as a control measure to ensure that any differences that might be found between training groups on the experimental task were not due to pre-existing differences between groups in pitch contour awareness, one of the most relevant factors for predicting success in tone learning (Wong & Perrachine, 2007). Results from this task showed similar accuracy rates across the three groups: Tone-focus: $M = 0.72$ (SD = 0.14), Vowel-focus: $M = 0.75$ (0.13) and Control: $M = 0.78$ (0.14). We conducted statistical analysis of these data using generalized linear mixed effect modeling with training group as fixed effect, and intercepts for participants and items as random effects. This and all subsequent statistical analyses were conducted in R (version, 3.6.0, R Core Team, 2019), using the lme4 package (version 1.1-21, Bates et al., 2015). We first dummy-coded training group with Control group as reference level. We then reran the same model with Tone-focus group as reference level. These models indicated no significant differences between the three training groups (all $p > .11$). Thus, we confirm that participants randomly assigned to the three training groups had overall similar pitch contour awareness before training.

3.2. Forced choice task

Except for one trial with a missing mouse click response, data from a total of 8099 trials (90 participants, 90 test trials) entered analysis. Figure 3 presents participants’ overall accuracy in selecting the named target by group, collapsing
over different trial types. Unexpectedly, the Control group shows the highest overall accuracy rate, while overall accuracy in the Tone-focus group appears numerically the lowest.

Figure 3. Overall accuracy by training group; error bars = 95% CIs

Of critical interest, however, was whether the different training groups would perform differently on different trial types, and more specifically, whether participants trained in a cue-focus condition would learn the focused cue better. Figure 4 illustrates accuracy by trial type for each group. We are especially interested in the three groups’ performance on Tone-pair and Vowel-pair trials, where only a single cue differentiated the two words, and participants in the two experimental groups were trained to allocate their attentional focus to one of those two cues respectively. Visual inspection indicates that for Tone-pair trials, the Tone-focus group performed better than the Vowel-focus group, but did not differ from the Control group. For Vowel-pair trials, we see the opposite pattern, with the Vowel-focus group showing higher accuracy than the Tone-focus group, but no difference from the Control group.

Figure 4. Overall accuracy by training group and trial type
For statistical analysis, the binomial accuracy data were submitted to a generalized linear mixed effect model with training group, trial type and their interaction as fixed effects, and intercepts for participants and items as random effects. In order to assess main effects, both training group and trial type were simple-coded, with Control group and Baseline trial as reference level, respectively. Table 3 presents the output of this model. With regard to main effects of training group, the significant negative estimate for Tone-focus (b = -0.62, p = .005) indicates overall lower accuracy in the Tone-focus compared to the Control group, while the difference between the Vowel-focus and the Control group was not significant (b = -0.34, p = .12). Turning to main effects of trial type, all three minimal-pair trials show significantly lower overall accuracy than baseline trials (p < .001). The largest negative estimate for Tone-pair (b = -1.89) confirms that words differing by tone alone present the overall greatest difficulty for word recognition. Only one interaction effect was significant in the overall model. To further explore this interaction as well as our predictions regarding the effects of cue-focus training, we conducted follow-up analyses on the data from each trial type separately.

**Table 3. Results of generalized linear mixed-effects model for accuracy.**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.77</td>
<td>0.14</td>
<td>12.88</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tone-focus</td>
<td>-0.62</td>
<td>0.22</td>
<td>-2.89</td>
<td>.005</td>
</tr>
<tr>
<td>Vowel-focus</td>
<td>-0.34</td>
<td>0.22</td>
<td>-1.57</td>
<td>.12</td>
</tr>
<tr>
<td>Consonant-pair</td>
<td>-0.86</td>
<td>0.10</td>
<td>-9.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tone-pair</td>
<td>-1.89</td>
<td>0.09</td>
<td>-21.53</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vowel-pair</td>
<td>-0.65</td>
<td>0.10</td>
<td>-6.60</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tone-focus:Consonant-pair</td>
<td>-0.01</td>
<td>0.24</td>
<td>-0.05</td>
<td>.96</td>
</tr>
<tr>
<td>Tone-focus:Tone-pair</td>
<td>-0.18</td>
<td>0.25</td>
<td>-0.70</td>
<td>0.48</td>
</tr>
<tr>
<td>Tone-focus:Vowel-pair</td>
<td>0.94</td>
<td>0.22</td>
<td>4.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vowel-focus:Consonant-pair</td>
<td>-0.28</td>
<td>0.22</td>
<td>-1.24</td>
<td>.21</td>
</tr>
<tr>
<td>Vowel-focus:Tone-pair</td>
<td>0.06</td>
<td>0.24</td>
<td>0.26</td>
<td>0.79</td>
</tr>
<tr>
<td>Vowel-focus:Vowel-pair</td>
<td>0.26</td>
<td>0.26</td>
<td>0.99</td>
<td>0.32</td>
</tr>
</tbody>
</table>

For Tone-pair trials, no significant differences were found between the Tone-focus group and the Control group (b = 0.08, p = .75), while the Vowel-focus group showed significantly lower accuracy than the Control group (b = -0.55, p = .03). For Vowel-pair trials, no significant differences were found between the Vowel-focus group and the Control group (b = 0.06, p = .86), while the Tone-focus group showed significantly lower accuracy than the Control group (b = -0.90, p = .006). These results suggest, contrary to the predictions of the assumption under investigation, that cue-focus training did not lead to more accurate learning of the focused cue, but to less accurate learning of non-focused...
cues. Furthermore, analyses of Baseline and Consonant-pair trials showed no significant differences between the Vowel-focus and the Control group (both $p > .14$), but the Tone-focus group had significantly lower accuracy than the Control group (both $p = .002$). This further indicates that focusing on tone, the novel and more difficult cue for these learners may have drawn their attention away from other, non-focused cues, with the consequence of negatively impacting their word learning overall.

4. Discussion

The goal of this study was to examine the effectiveness of cue-focus training in word learning. More specifically, we wanted to test an assumption implicit in previous research and pedagogical practice on the learning of lexical tones, namely that manipulation of learners' attentional focus to the contrast of tonal cues would benefit their learning of words with lexical tones. Since many factors are likely to influence learning outcomes in more naturalistic learning settings, we aimed to minimize the role of previous language experience by training English speakers in three different training groups to learn novel words in a laboratory setting. Learning outcomes were measured by the accuracy rate in a forced-choice spoken word recognition task following training. If tone-focus training is effective, we expected to see higher accuracy in the Tone-focus than in the Control group, especially in trials where tone was the only cue differentiating the two words (Tone-pair trials). If cue-focus training is effective more generally, we would expect to see an analogous pattern with participants trained in the Vowel-focus group outperforming the Control group, especially in trials where vowel quality was the only cue differentiating the two words (Vowel-pair trials).

Unexpectedly, we found that participants trained in the Tone-focus group showed overall lower accuracy in word recognition than those in the Control and Vowel-focus groups. More importantly, even in Tone-pair trials, where tone was the only cue differentiating the two words, the Tone-focus group achieved similar accuracy as the Control group, indicating tone-focus training was not effective. Critically, the Vowel-focus group performed significantly lower than the Control group on Tone-pair trials, suggesting Vowel-focus training hurt learning of tone. This is further supported by the fact that this pattern was mirrored with Vowel-pair trials, where no difference was observed between the Control and Vowel-focus groups, but the Tone-focus group showed significantly lower accuracy than the Control group.

In sum, and contrary to our initial predictions, the findings from this study suggest that cue-focus training did not improve learning of the focused cue, but instead hurt learning of non-focused cues. Since tonal words are composed of both tones and segments, training focusing on a single cue such as tone might thus hurt the learning of a word as a unit. These results are consistent with Zhao et al.’s (2011) hypothesis that “the recognition of Chinese monosyllabic words might rely more on global similarity of the whole syllable structure or syllable-
based holistic processing rather than phonemic segment-based processing” (p. 1761). If learners’ attention is fragmented by considering each cue separately during learning, this might prevent them from processing words holistically during later word recognition.

Most previous studies investigating the acquisition of lexical tone tested learning outcomes through tone judgment tasks, where learners only need to decide whether two auditorily presented syllables carry the same tones or not (Liu et al., 2011), or to match auditorily presented syllables by tone (Godfroid et al., 2017). This type of test focuses uniquely on identification of the form of a single cue. In our study, we used a word recognition task, where listeners needed to identify the form of both segments and tones and use them together to access lexical meaning in order to select a visually depicted referent. This presents a considerably more demanding task.

We chose to focus on spoken word recognition based on the consideration that the fundamental function of tone is as a component in distinguishing between words. Bearing this in mind, any teaching or learning methods that do not work towards the successful use of lexical tones in word learning would seem misguided, and methods not able to increase the accurate processing of tones in the context word recognition would ultimately be ineffective. The current study shows that cue-focus training is not always effective. In our laboratory-based auditory novel word learning experiment, neither the Tone-focus nor the Vowel-focus group showed higher accuracy than the Control group in overall word recognition. However, the cue-focus training did reallocate participants' attentional focus, yet not to the benefit of the focused cue, but to the detriment of non-focused cues.

The results from this study do not necessarily conflict with findings from previous research that suggested certain versions of tone-focus training (e.g. contour + pinyin condition in Liu et al., 2011) might be more effective than others (e.g. contour only condition) at reducing errors in tone identification. Yet they highlight the importance of considering the difference between tone identification in isolated syllables compared to meaning-bearing words, and they indicate that it may be helpful to include a control group receiving no cue-focus training for comparison.

The current findings also do not directly conflict with existing learning theories. As the Noticing hypothesis (Schmidt, 1990) states, noticing is necessary, but it is not necessarily sufficient for intake. Consistent with the Automatic Selective Perception model (Strange, 2011), influenced by their native experience of English, our participants paid less attention to tone, showing lowest accuracy in Tone-pair trials across the three training groups, while the experimental training conditions did reallocate their attentional focus, but at the expense of paying less attention to non-focused cues. The Competition Model (MacWhinney, 2005, 2012) suggests that presenting the contrastive form can increase the relative strength of the cue in acquisition. However, we did not observe a benefit for word learning by presenting the contrastiveness of a cue. This might be owing to how we measured the learning outcome. Instead of
attention to form alone as in previous studies, our word learning and recognition tasks required use of both form and meaning. Our findings suggest that in the context of word learning, where both tones and segments must be taken into consideration, trying to increase the relative cue-strength of tones by highlighting the contrastiveness of this cue in training is not effective, and may ultimately hurt syllable-based holistic processing and learning of words in tonal languages.

5. Future directions and Conclusion

Our analyses of learning outcomes in this paper were based on participants' final choices in a two-alternative forced-choice task. This is a relatively crude measure that may mask subtler differences between groups in their processing of the auditory input prior to their final mouse click. Eye-movements to potential referents in a visual scene can reveal more subtle aspects of lexical activation during real-time spoken word recognition (e.g. Allopenna et al., 1998; Quam & Creel, 2017). Although we have only reported final choice (mouse click accuracy) data here, eye-gaze data was collected during the forced-choice task. This data is currently under analysis. In particular, we plan to analyze data from trials with correct responses to examine whether participants trained in different groups showed different patterns of fixation to the referents in the visual scene prior to making a final choice.

In addition to the procedure described here, all participants also completed a production task after the forced-choice task, in which they were asked to name the 18 target referents. Considering training occurred only in a single session, this task was expected to be difficult for participants, and was included only for exploratory purposes. Analysis of these production data, with a focus on potential differences between the three training groups, is currently in progress.

As pointed out above, the current study tested participant's learning outcomes in a task that required lexical access, thus involving both meaning and form. In future work, it might be interesting to add a judgment task to see to what extent the between-group differences observed in our word learning task would extend to a task requiring attention to form only.

In conclusion, this study has provided at least some evidence indicating the commonly held assumption that cue-focus training is effective for learning lexical tones does not always hold. In particular, we found that cue-focus training did not benefit the learning of the focused cue, but instead hurt the learning of non-focused cues, which in turn led to a decrease in overall word learning success. Thus, the results from the current experiment suggest that vocabulary learning in a tonal language may be better supported through syllable-based holistic training than through allocating attentional focus on a specific phonemic cue. More studies need to be done to test this hypothesis.
References


Ling, Wenyi, & Grüter, Theres (2018, November). Lexical tone in L2 Mandarin: The relation between categorical perception and real-time spoken word recognition. Poster presented at the BUCLD 43, Boston, MA.


