



Cantonese tone word learning by tone and non-tone language speakers

Angela Cooper and Yue Wang

Department of Linguistics, Simon Fraser University, Burnaby, B.C., Canada

akcooper@sfu.ca

Abstract

Adult non-native perception is subject to influence from a variety of factors, including native language experience. The present research examines the effect of linguistic experience on non-native tone perception and tone word learning. Native Thai and English-speaking participants completed seven sessions of lexical identification training on words distinguished by Cantonese tones. A tone identification task was administered before and after training. Both groups had comparable tone identification accuracy; however, Thai listeners obtained greater tone word learning proficiency. The findings suggest that native language experience with employing pitch lexically facilitates the incorporation of non-native tones into novel lexical representations.

Index Terms: lexical tone, word learning, Cantonese, non-native perception, linguistic experience

1. Introduction

1.1. Background

Over the course of native language (L1) development, the perceptual sensitivities of a given speaker become attuned to the critical acoustic characteristics of their L1, which may later cause “perceptual interference” when attempting to tune into the important cues of a foreign language [1]. However, it is not the case that all contrasts are uniformly challenging for all listener groups [2]. Learners’ L1 phonetic systems can interact with the second language (L2) system, shaping perception and the formation of new phonetic categories (e.g. [3]).

Previous studies have reported that listeners’ native phonetic systems affect the perception of foreign sounds, and that the interaction of new phonetic structures, both segmental and suprasegmental, with native ones can have a significant influence on learning [4]. Regarding lexical tone, research has suggested that having experience with being attuned to native pitch distinctions can facilitate the acquisition of non-native tones [3, 5]. On the other hand, other studies have suggested that native tone language experience does not necessarily predict greater success over those with non-tone language backgrounds in perceiving non-native tones. For example, [4] found no significant difference in Cantonese tone identification accuracy by English and Mandarin listeners. Instead, they proposed that the nature of the native tonal inventories and how they interact with incoming non-native tones may better elucidate differences in cross-linguistic tone perception. For instance, Mandarin listeners’ identification accuracy was best on the three Cantonese lexical tones that have similar counterparts in Mandarin, suggesting that native category representations can have a significant influence on non-native perception. Additionally, it has been reported that perceptual discrepancies in lexical tone perception can be attributed to language-specific weightings of two perceptual dimensions: pitch height and direction of change [6]. It was posited that tone and non-tone language groups are separable based on their respective rankings of these perceptual

dimensions, as ‘direction of change’ appeared to be the most perceptually salient dimension for Thai listeners; whereas, non-tone language listeners (English) gave greater weight to the ‘height’ dimension.

While it is evident that L1 experience influences the ease with which some phonemic contrasts are acquired, the majority of previous research has focused on its role in listeners’ perceptual learning of individual phonetic distinctions. There is a paucity of research investigating the impact of L1 on the ability to apply non-native contrasts to broader linguistic contexts such as word learning. In a study training French and English listeners to utilize the Thai three-way voicing and aspiration contrast to distinguish lexical items, it was reported that participants were successful at using this contrast for lexical purposes, even after a relatively short period of training [7]. The authors concluded that listeners initially form lexical representations utilizing L2 features contrastive in their L1 (in this case, voicing), before lexicalizing features that are not present or phonemically relevant in their L1 phonology. [8] extended these results to suprasegmentals, reporting that English listeners were capable of using non-native lexical tones to distinguish word meaning.

However, little research has compared the influence of tonal versus non-tonal native language background on learning non-native words minimally distinguished by tone.

1.2. The current study

The present research examines the role of linguistic experience on Cantonese tone word learning using non-native listeners with tonal (Thai) and non-tonal (English) backgrounds. Native Thai and English listeners completed a tone word training program where they learned the meanings of 15 novel vocabulary words, minimally distinguished by five Cantonese tones. A Cantonese lexical tone identification task was also administered before and after training. This study seeks to extend research on the influence of L1 experience (tone vs. non-tone language) on non-native tone perception to investigate its effect on the ability to utilize non-native tones in a linguistic context, namely to distinguish lexical items.

Although previous findings have shown that linguistic experience can facilitate [3, 5] or inhibit tone perception [4], we hypothesize that the Thai group will attain greater proficiency at tone word identification than the English group. This hypothesis is motivated by the fact that the training is a linguistic task, and Thai listeners will have more experience utilizing pitch distinctions to make lexical differentiations. Concerning the lexical tone identification task, the identification of specific L2 tones may be facilitated by similar L1 tone contours [4], such as high or rising tones for English listeners (given L1 intonational patterns) or the level and falling tones for the Thai listeners.

2. Methods

2.1. Participants

Eighteen native Thai and sixteen native Canadian English adults participated in this study. All participants had no prior knowledge of Cantonese or any other lexical tone language (other than their L1). They also had less than 4 years of musical experience, and no experience within the last 5 years. Furthermore, they were college-educated and possessed normal hearing and cognitive abilities. The Thai group (10 male, 8 female; mean age: 22 years) was recruited from Chulalongkorn University in Bangkok, Thailand, and considered the Bangkok dialect (Standard Thai) to be their first and dominant language. The English participants (6 male, 10 female; mean age: 24) were recruited at Simon Fraser University and the University of British Columbia in Vancouver, Canada.

2.2. Stimuli

2.2.1. Pre-/post training identification task

Two native Cantonese speakers (1 male, 1 female) produced five CV monosyllables (*waj*, *low*, *si*, *pej*, *fu*) with five Cantonese tones (high-level, high-rising, low-falling, low-rising and low-level), for a total of 25 real-word stimuli. The mid level-tone was not included, as it may be easily confused for the high and low level tones, particularly in the absence of any contextual cues [4]. To maintain focus on the suprasegmental information, the phonemes were common to Thai, English and Cantonese.

2.2.2. Training

Four novel speakers (2 male, 2 female) not used in the pre-/post-tests produced three CV monosyllables (*tsou*, *k^vaaj*, *wu*) with five Cantonese tones. These 15 words were associated with meanings (common concrete nouns), as represented by a picture presented on the screen. Pictures were selected from a set of 260 standardized pictures, controlled for visual complexity and cultural familiarity [9]. Because participants would be receiving lexical identification training (learning sound-meaning pairings), these particular syllables were selected because they do not contain semantic content in Thai or English, so as to reduce lexical competition with existing words in participants' lexicons. The speakers were recorded in a sound-attenuated booth at a 44.1 kHz sampling rate.

2.3. Procedure

2.3.1. Pre-/post-training identification task

Both groups first completed a familiarization task in order to become familiar with the five Cantonese tones and learn how to identify them. They heard each Cantonese tone pronounced in isolation and viewed a corresponding tone diagram on the screen. The participants were then asked to respond after each stimulus, identifying the tone they heard by pressing the number on the keyboard corresponding to the appropriate tone diagram. They received feedback on the accuracy of their response as well as the correct answer. This task used productions of /ji/ by the female pre-/post-test talker. Three randomized repetitions produced a total of 15 trials, lasting approximately 2 minutes.

The main task was a five alternative forced-choice identification task, where the participants identified the tone of each syllable, similar to the familiarization section. However,

they did not receive any feedback on their identification accuracy. They identified 100 randomized stimuli (5 syllables x 5 tones x 2 speakers x 2 repetitions), presented with an inter-stimulus-interval of 3 seconds. The task took approximately 8 minutes.

2.3.2. Training

Participants engaged in a series of seven training sessions, approximately 30 minutes each, administered on four separate days over the course of two weeks. Each day of training was separated by 2-4 days. They learned the full set of 15 training words and their associated meanings in each session. Training listeners on sound-meaning associations was utilized to simulate a more "natural" learning paradigm. Stimulus presentation and testing procedures were modeled after training in [7-8].

Each training session included 5 training blocks, 2 review blocks and a session test. The format of each training block consisted of listening to 4 randomized repetitions of 3 words while viewing the visual representation of their meaning (2 speakers x 3 words x 2 repetitions). Each block of words contained non-minimal triplets (i.e. three different tones on three different syllables). Every block concluded with a small quiz on the three words learned in that block, whereby participants heard a stimulus and were presented with the three pictures of the words they had just learned (4 speakers x 3 words). They needed to indicate the correct meaning for the word by selecting the appropriate picture. They were provided with feedback on their accuracy, as well as the correct answer and a re-playing of the stimulus. Participants then completed two additional blocks to review the training items. The first review was comprised of all 15 words, blocked by syllable, produced by one female speaker from training (15 trials). Blocking for syllable allowed participants to hear the minimal tone quintuplets and highlight the tonal distinctions. The second review consisted of all 15 words produced by 2 speakers (1 male, 1 female), where participants chose the meaning after each stimulus by selecting the appropriate picture from all 15 options on the screen (30 trials). Both review blocks provided feedback after each response, similar to the block quizzes. Each session concluded with a test where participants identified all 15 words learned in the training program without feedback. The session test followed the same format as the final review block, now involving the 15 training words produced by all 4 speakers (60 trials). A training session lasted approximately 20-25 minutes. Each participant received 2 training sessions per day (except for the last day of training where they received only one), with a 15-minute break between each training session.

For all tasks, comparable settings were ensured for the Thai and English participants by using the same headphone sets, volume levels and PC laptops and for testing. Task instructions and feedback information were provided in Thai and English for the respective participant groups.

3. Results

3.1. Pre-/post-training tone identification

Identification accuracy on the pre- and post-training tasks was tabulated based on the proportion of correct responses by lexical tone. The mean percent correct scores were submitted to a 3-way mixed analysis of variance (ANOVA) with Group (English, Thai) as a between-subjects factor and Test (pre, post) and Tone (high-level, high-rising, low-falling, low-rising, low-level) as repeated measures.

A significant main effect of Test was obtained [$F(1,32)=19.40, p<.0001$], indicating that participants significantly increased their lexical tone identification accuracy after training (41% to 54%). However, no significant group difference was found in tone identification accuracy across tests [$F(1,32)=2.505, p=.123$], as English (51%) and Thai listeners (45%) performed comparably. The interaction of Test x Group was also not significant [$F(2,32)=.965, p=.392$].

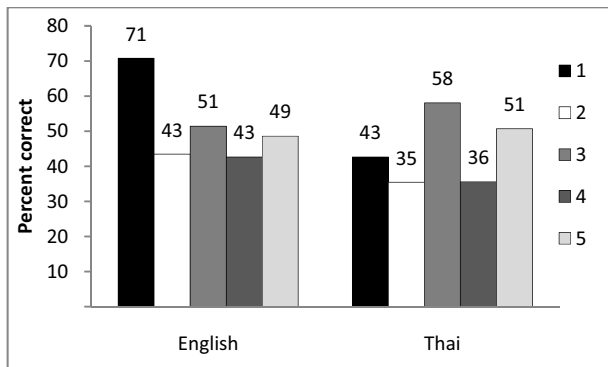


Figure 1: Mean identification accuracy by tone across pre- and post-tests by English and Thai listeners. Tones: 1=high level, 2=high-rising, 3=low-falling, 4=low-rising, 5=low-level.

The ANOVA yielded a significant main effect of Tone [$F(4,32)=18.239, p<.0001$] and Tone x Group interaction [$F(4,32)=11.829, p<.0001$]. Additional 1-way ANOVAs for each group with Tone as the independent variable revealed significant differences in tone identification accuracy patterns (Figure 1). For the English group, pairwise comparisons indicated that high-level tones were the easiest to identify across tests, significantly better than all other tones ($p<.001$). No significant difference in accuracy was found between high- and low-rising and low-level tones ($p>.574$). The Thai group's highest identification scores were for low-falling tones, significantly greater than high-level, high- and low-rising tones ($p<.045$). Low-level tones were also identified with greater accuracy than low and high-rising tones overall ($p<.033$). No significant Test x Tone x Group interaction was obtained [$F(4,32)=.935, p=.446$].

Tonal confusion analyses were also conducted, whereby for a given tone (heard stimulus), the proportions of misidentifications were tabulated by tone (response given). These sets of confusions (4 for each tone) were each submitted to 3-way ANOVAs with Group as a between subjects factor and Test and Confusion pattern as repeated measures. For brevity, only results with a significant Group x Confusion pattern interaction are reported here. For high-level tones, Thai listeners misidentified them as high- or low-rising significantly more than other tones ($p<.001$), but the English group had no significant difference between tonal confusions ($p<1.00$). Low-falling tones were most frequently confused as low-level by English listeners ($p<.0001$), but more commonly as low-rising ($p=.006$) and low-level ($p<.008$) by Thais. However, no significant difference was found between the misidentifications of low-level as low-rising and high-rising for the Thai group. Low-rising tones were confused most often as high-rising or low-level by the English group ($p<.013$); whereas, confusion patterns were more evenly distributed for the Thais, with no significant difference between the confusions ($p=1.00$).

3.2. Tone word training

To evaluate overall improvement in tone word identification over the course of training, mean percent correct was calculated for the first and last training session tests (Figure 2). A 3-way mixed ANOVA was conducted, with Group as a between-subjects factor, and Session (1, 7) and Tone as repeated measures.

A significant main effect of Session was found, as overall tone word identification accuracy increased from the first to last session across groups (22% to 62%) [$F(1,32)=237.109, p<.0001$]. The ANOVA also yielded a significant Session x Group interaction [$F(1,32)=8.259, p=.007$]. Subsequent 1-way ANOVAs for each session with Group as the independent variable were conducted. No significant group differences were found for the first session [$F(1,32)=.075, p=.786$]. However, significance was achieved on the last session [$F(1,32)=7.031, p=.012$], indicating that the Thai listeners achieved higher overall tone word identification accuracy (71%) than the English listeners (54%) by the last session.

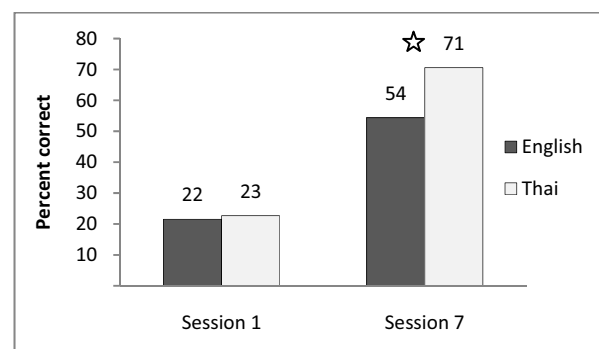


Figure 2: Mean word identification accuracy for first and last training sessions by English and Thai groups. “*”: statistically significant difference ($p<0.05$).

A significant Tone x Group interaction was also obtained [$F(4,32)=3.966, p=.005$]. Further analyses on each tone revealed that Thai listeners significantly outperformed the English group on high-rising (50% vs. 32%, $p=.004$) and low-level (45% vs. 31%, $p=.040$) tone word items across tests. No significant Session x Tone x Group interaction was found [$F(4,32)=2.042, p=.092$].

4. Discussion

Our hypothesis that Thai listeners would attain higher tone word learning proficiency than English listeners was confirmed, as Thais had greater tone word identification accuracy by the end of training, pointing to an effect of L1 background. These results suggest that Thai listeners' native language experience with using pitch to differentiate word meaning can beneficially transfer to aid acquisition of non-native tone words. This is consistent with [7], who posited that listeners first construct lexical representations utilizing features that are lexical in their native language before features that are not present or phonemically relevant in their L1. This may explain why Thai listeners, whose L1 uses pitch lexically, were able to achieve higher word learning success than the English listeners.

These findings could be considered with respect to the native language functional load of pitch [10-11]. The functional load of a given contrast can be defined in terms of the frequency of its occurrence as well as the level of contrastivity [12], that is, how many minimal pairs exist in the

language, as well as to what degree these pairs are contrastive (e.g. lexical, pragmatic). In English, pitch has relatively low functional load, as stress and intonation are primarily used to mark grammatical contrasts or denote pragmatic or emotive information. On the other hand, pitch in Cantonese and Thai has high functional load, as it is used phonemically on all words. Given the present findings, it is perhaps more challenging for listeners to acquire words where there is an L1-L2 disparity in functional load for a contrast, particularly when they are required to shift from low to high functional load [13]. These listeners need to learn not only to attune to cues that hold less linguistic significance in their native language but also to apply them to make lexical contrasts.

On the other hand, with regards to the pre-/post-training tone identification task, it appears that having a tone language background is not necessarily advantageous for non-native phonemic tone perception, consistent with previous findings reporting no significant advantage of L1 tone experience on the identification of L2 tones [4]. The present results showed no significant difference in tone identification accuracy between the Thai and English groups across tests. However, group differences did arise with respect to tonal confusion and tonal accuracy patterns, which could be attributed to L1 influence. For instance, English listeners most commonly misidentified low-falling tones as low-level; whereas, the Thai group's misidentifications were split between low and high-rising and low-level tones. These differences in confusion patterns may be derived from a discrepancy in the weighting of perceptual dimensions in the L1. [6] suggested that non-tone language listeners attend to pitch height to a greater degree than direction, which may explain the English group's tendency to identify low-falling as low-level. Conversely, Thai listeners' confusing low-falling as low or high-rising suggests they are attending to the changing contour direction (regardless of the actual direction of change) more than height. [4] reported a similar result for Mandarin listeners, who also initially misidentified low-falling tones as low and high-rising tones. Furthermore, English listeners were most successful at identifying the high-level tone, which could be attributed to a successful mapping of this L2 tone to the high tone in the English intonational system (which is considered to be comprised of high and low pitch accent sequences [14]). Thai listeners, on the other hand, had the most success with low-falling and low-level tones, which may have resulted from the Thais assimilating the low-falling tone to their L1 Falling category and the Cantonese low-level tone to their L1 Low category. This may have facilitated their performance on these two tones. These results illustrate that the L1 phonetic inventory may have a substantive influence on non-native tone perception.

It appears that different aspects of language background are drawn upon in different domains of acquisition (e.g. phoneme categorization versus lexical learning). L1 categories and native weightings of perceptual dimensions influence the categorization of L2 phones, which, depending on the specific nature of the phonetic L1 and L2 inventories, may or may not be facilitative [4]. In this study, Thai and English listeners performed comparably on the pre-/post-training tone identification task, regardless of the function of pitch cues in their native languages. However, in the word learning domain, experience with its function as being lexically contrastive in Thai became advantageous for the Thai group in acquiring non-native lexical items, presumably because Thais had more experience with incorporating pitch information into lexical representations.

5. Conclusions

The present study examined the effect of linguistic experience on non-native phonemic tone and tone word identification. The findings suggest that prior tone language experience does not necessarily aid non-native tone identification, as no significant difference was found between Thai and English listeners in tone identification accuracy. Instead, the nature of their respective L1 tonal inventories and perceptual cue weightings may provide some insight into group performance discrepancies on specific tonal contrasts. However, L1 experience with using tone phonemically appears to facilitate the use of non-native pitch contrasts to distinguish novel lexical items. The current research thus extends the previous work on the role of L1 influence to tone word learning, with different L1 backgrounds (tonal or non-tonal) and at different linguistic domains (phonemic versus lexical).

6. Acknowledgments

The authors would like to thank Pattama Visuttipitukul, Tachai Luangvaranunt, Akkaporn Cooper, Amy Lee and David Potter for their assistance and support. This research is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (312457-2006).

7. References

- [1] Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C., "A perceptual interference account of acquisition difficulties for non-native phonemes", *Cognition*, 87(1):B47-B57, 2003.
- [2] Flege, J. E., Bohn, O., and Jang, S., "The effect of experience on nonnative subjects' production and perception of English vowels", *Journal of Phonetics*, 25: 437-470, 1997.
- [3] Lee, Y-S., Vakoch, D. A. and Wurm, L. H., "Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison", *Journal of Psycholinguistic Research*, 25:527-542, 1996.
- [4] Francis, A., Ciocca, V., Ma, L. and Fenn, K., "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers", *Journal of Phonetics*, 36:268-294, 2008.
- [5] Wayland, R. P. and Guion, S. G., "Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report", *Language Learning*, 54:681-712, 2004.
- [6] Gandour, J., "Tone perception in Far Eastern languages", *Journal of Phonetics*, 11:149-175, 1983.
- [7] Curtin, S., Goad, H., and Pater, J. V., "Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers", *Second Language Research*, 14(4):389-405, 1998.
- [8] Wong, P.C. and Perrachione, T. K., "Learning pitch patterns in lexical identification by native English-speaking adults", *Applied Psycholinguistics*, 28:565-585, 2007.
- [9] Snodgrass, J. G., and Vanderwart, M., "A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity", *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):174-215, 1980.
- [10] Gandour, J. T., Wong, D., Lowe, M., and Dziedzic, M., "A Cross-Linguistic fMRI Study of Spectral and Temporal Cues Underlying Phonological Processing", *Journal of Cognitive Neuroscience*, 14(7):1076-1087, 2002.
- [11] Van Lancker, D., and Fromkin, V. A., "Hemispheric specialization for pitch and "tone": Evidence from Thai", *Journal of Phonetics*, 1:101-109, 1973.
- [12] King, R. D. "Functional Load and Sound Change", *Language*, 43(4):831-852, 1967.
- [13] McAllister, R., Flege, J. E., and Piske, T., "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian", *Journal of Phonetics*, 30(2):229-258, 2002.
- [14] Beckman, M. E., and Pierrehumbert, J., "Intonational structure in Japanese and English", *Phonology Yearbook*, 3:255-309, 1986.