EFFECT OF SEMANTIC CONTEXT ON THE PERCEPTUAL LEARNING OF LEXICAL TONE

Yue Wang[a], *Angela Cooper*[a], *Xianghua Wu*[a] *& Dawn Behne*[b]

[a]Department of Linguistics, Simon Fraser University, Canada;
[b]Department of Psychology, Norwegian University of Science and Technology, Norway
yuew@sfu.ca

ABSTRACT

Previous research has not determined whether the inclusion of lexical semantic information facilitates or inhibits the learning of second language (L2) phonetic contrasts. The present study addresses this issue by comparing the acquisition of Mandarin Chinese tones with and without semantic contexts. Two groups of native English listeners with no lexical tone experience participated in a Mandarin tone training program where one ("No meaning") group received training with only phonetic tonal contrasts, while the other ("Meaning") group was additionally provided with semantic information. Results show that although both groups started comparably and improved significantly with training, the "No-meaning" trainees had significantly higher tone identification accuracy rates than the "Meaning" trainees after training. However, the inter-session tests with the training stimuli reveal the opposite pattern, where the Meaning group outperformed the No-meaning group. Together, these results indicate that, at the initial stage of tone learning, non-native listeners learn more efficiently by focusing on phonetic tonal distinctions, whereas remembering the meanings of tone words does not generalize well to tone category identification.

Keywords: semantic information, lexical tone learning, non-native speech perception, Mandarin

1. INTRODUCTION

Over the course of native language (L1) development, acoustic information relevant to the L1 is weighted more heavily than less relevant information, making it a significant challenge for adult learners of a foreign language to re-attune their perceptual systems to the appropriate acoustic cues to discern non-native phonetic distinctions. Laboratory training has nevertheless been found to improve listeners' perception of non-native segmental and suprasegmental contrasts, demonstrating that adult perceptual systems retain a degree of plasticity [6, 8].

Given that listeners' perceptual mechanisms for categorizing sounds can be shaped by perceptual training, research has also investigated a range of training procedures to determine how to maximally improve the perceptual accuracy of non-native sounds. Such methods have included high-variability [6, 8], audio-visual [4] and sentence versus word context training [5]. Additionally, some studies have suggested that providing word meaning can mediate acquisition of phonetic contrasts [1, 3]. For example, the discrimination of a novel phonemic distinction was improved by the addition of lexical information as compared to learners who received only auditory information [3]. It has been posited that providing meaning may create a larger incentive for listeners to focus on the subtle acoustic distinctions between the words and extract the relevant cues necessary to distinguish them [1].

However, other studies have indicated that attentional mechanisms play a role in phonetic learning [2], such that orienting the listeners' attention away from the appropriate phonetic information can inhibit the acquisition of non-native contrasts. For example, a group of English listeners who were asked to attend to semantic information had poorer discrimination performance with nonnative Hindi stop contrasts than a group who focused on the phonetic information [2]. Additionally, the presence of lexical information demands mapping sound to meaning, resulting in increased cognitive load and thus interference with the learning of phonetic details [9].

Based on these findings, it is not clear whether the inclusion of lexical information during phonetic training facilitates or inhibits the learning of non-native contrasts. Furthermore, research has not examined the influence of lexical information on acquiring non-native suprasegmentals such as lexical tone. Thus, the present study intends to address these issues by providing two types of perceptual training on the identification of Mandarin lexical tones for native English listeners. This was accomplished by comparing the overall improvement in tone identification by listeners who received only phonetic training (i.e., "No-meaning" group) against those who received lexical content in addition to phonetic information during training (i.e., "Meaning" group).

2. METHODS

2.1. Participants

Twenty-six native Canadian English adults with no prior knowledge of Mandarin or any other lexical tone language participated in this study. They had normal hearing and no musical or pitch-related training experience (*M*=0.96 years). Fourteen participants were randomly assigned to the No-meaning (NM) group (*M age*=23; 5 male, 9 female), and twelve were included in the Meaning (M) group (*M age*=21, 2 male, 10 female).

2.2. Stimuli

2.2.1. Pre-/post training tone identification

Two native Mandarin speakers (1 male, 1 female) produced twelve monosyllables (*zhuo, xiong, run, zi, que, chi, ka, pou, fu, lan, nin, ting*) with four Mandarin tones (high-level, rising, dipping, falling), for a total of 48 audio stimuli.

2.2.2. Training

Two novel speakers (1 male, 1 female) not used in the pre-/post-tests produced six monosyllables (*ri, chun, qiong, xue, cuo, zhi*) with four Mandarin tones, for a total of 24 stimuli. For the Meaning group, these words were assigned meanings (common concrete nouns) represented by pictures, which were selected from a set of 260 standardized pictures, controlled for visual complexity and cultural familiarity [7].

2.3. Procedure

2.3.1. Pre-/post-training tone identification task

A familiarization task was administered to allow participants to become familiar with the four Mandarin tones and learn how to identify them. They heard each Mandarin tone pronounced in isolation and viewed a corresponding tone diagram on the screen. Next, the participants were asked to identify the tone after each stimulus 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 */fa/* by the female pre-/post-test talker. Four randomized repetitions produced a total of 12 trials, lasting approximately 2 minutes.

For the main task, both groups of participants completed a four alternative forced-choice identification task, where they identified the tone of each syllable, similar to the familiarization section, but with no feedback. They identified 96 randomized stimuli (12 syllables x 4 tones x 2 speakers), presented with an inter-stimulus-interval of 3 seconds. The task took approximately 10 minutes.

2.3.2. Training

Participants completed six training sessions of 25 minutes each, administered on three separate days over the course of ten days. Each training day consisted of two sessions followed by a test. Both sessions began with a brief overview of the tones, where listeners would hear each tone in isolation and view its associated tone diagram. Similar to the familiarization task, training involved the participants responding after each stimulus by identifying the tone they heard and receiving feedback on the accuracy of their responses. The six training syllables were divided evenly between the two sessions. Each session contained eight randomized repetitions of three different syllables for a total of 96 trials (3 syllables x 4 tones x 2 speakers x 8 repetitions), presented with an ISI of 2 seconds. For participants in the M group, the assigned meaning for each of the words was presented on the screen while the stimulus was played. For the NM group, the screen was blank during stimulus presentation.

After finishing both sessions, all participants completed a test of the same format as their training sessions, but with no feedback. They were tested on 96 randomized trials involving all six training syllables (6 syllables x 4 tones x 2 speakers x 2 repetitions).

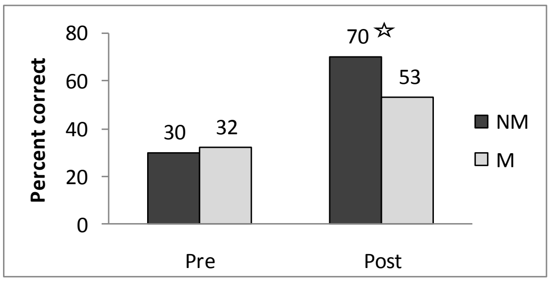
Participants completed these tasks in a sound-attenuated booth on PC computers wearing AKG K1441 Studio headphones using a comfortable listening volume.

3. RESULTS

3.1. Pre-/post-training tone identification

Identification accuracy on the pre- and post-training tests was calculated based on the proportion of correct responses by lexical tone (Figure 1). The mean percent correct scores were submitted to a 3-way mixed analysis of variance (ANOVA) with Group (NM, M) as a between-subjects factor and Session (pre, post) and Tone (1-4) as repeated measures.

Figure 1: Mean identification accuracy (%) for pre/post-tests by No-meaning (NM) and Meaning (M) groups. "*": statistically significant difference (p<.05).



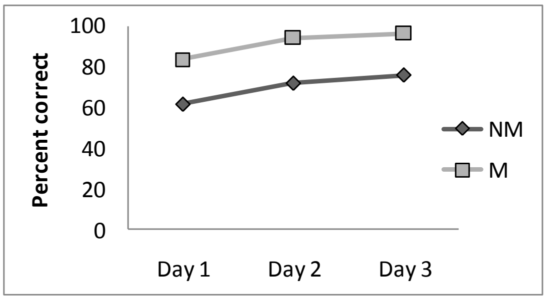
A significant main effect of Session was obtained [F(1,24)=96.691, *p*<0.0001], indicating that across groups, there was a significant increase in lexical tone identification accuracy after training (31% to 62%). The ANOVA also yielded a significant Session x Group interaction [F(1,24)=5.465, *p*=.028]. One-way ANOVAs for each group with Session as repeated measures obtained significance for both NM [F(1,13)=90.269, *p*<0.0001] and M groups [F(1,11)=23.057, *p*=.001], demonstrating that both groups made significant improvements as a result of training. Subsequent 1-way ANOVAs for each session with Group as the independent variable were conducted. No significant group differences were found for the pre-test [F(1,24)=.126, *p*=.725]; however, there was a significant group difference in the post-test [F(1,24)=5.591, *p*=.026], indicating that listeners in the NM group (70%) had significantly higher accuracy rates than those in the M group (53%) after training.

Moreover, a significant effect of Tone was observed [F(3,22)=23.6, p<.0001], along with a significant Tone and Session interaction [F(8,18)=19.9, p<.0001]. Subsequent one-way ANOVAs for each session with Bonferroni-adjusted post hoc analysis revealed that, across groups, while Tone 2 accuracy was comparable with that of the other tones at pre-test, it was the lowest of all the tones at post-test, indicating that the degree of improvement resulting from training was the smallest for Tone 2.

3.2. Training

In order to examine how the participants were progressing throughout training, the mean tone identification accuracy was tabulated for each test at the conclusion of each training day (i.e., inter-session test). A 3-way ANOVA with Group as a between-subjects factor, and Day (1-3) and Tone (1-4) as repeated measures revealed a significant main effect of Day [F(2,24)=39.217, *p*<0.0001], and Bonferroni-adjusted pairwise comparisons indicated that identification accuracy across groups significantly increased on each successive training day (*p*<.007). Furthermore, a significant main effect of Group was also found [F(1,24)=22.037, *p*<0.0001], as the M group had significantly higher tonal accuracy scores (91%) than the NM group (70%) across training day tests. Day x Group [F(2,24)=.141, *p*=.869] and Day x Tone x Group [F(6,24)=1.906, *p*=.084] interactions did not reach significance.

Figure 2: Mean identification accuracy for inter-session tests for No-meaning (NM) and Meaning (M) groups.



The ANOVA also yielded a significant interaction of Tone x Group [F(3, 24)=8.104, *p*<0.0001]. Additional 1-way ANOVAs on each tone with Group as the independent variable revealed that the M group was significantly more accurate than the NM group on each tone across training day tests (*p*<.007). For the NM group, Tone [F(3,13)=14.629, *p*<0.0001] had a significant main effect on tonal accuracy. Pairwise comparisons (Bonferroni) indicated that identification of Tone 2 was significantly worse than all other tones (*p*=.002). No significant differences between tonal accuracy were found for the M group [F(3,11)=.216, *p*=.651].

4. DISCUSSION AND CONCLUSIONS

Results show that both the Meaning and No-meaning groups made significant improvements as a result of training. However, while the two groups started at the same level at pre-test, the No-meaning trainees had significantly higher tone identification accuracy rates than the Meaning trainees at post-test. In contrast, the inter-session tests with the training stimuli reveal that the Meaning group outperformed the No-meaning group throughout training.

These results indicate that, for the Meaning group in contrast to the No-meaning group, the degree of improvement during training did not proportionally generalize to the identification of the tonal syllables not used in training. While the Meaning group's inter-session test scores almost reached ceiling (averaging 91%), their post-test tone identification accuracy was only 53%. As such, these trainees presumably learned to associate the whole entity of each training stimulus (i.e., cumulative segmental, tonal, and lexical semantic information) to the corresponding word (represented as an object), rather than focusing on the tonal contrasts per se. These patterns were further demonstrated by the results of individual tones. That is, Tone 2 accuracy was lower than the other tones at post-test for both groups. However, only for the No-meaning group was Tone 2 perception consistently low during training, while the Meaning group's inter-session performance was equally good across tones, indicating that the latter group focused on word rather than tone during training.

The current findings of tone learning do not support the previous claim from segmental learning research that the inclusion of word meaning may enhance focus on the subtle acoustic distinctions between words [1, 3]. We argue that, at the initial stage of learning non-native tonal contrasts, training with the single dimension of tonal information can be more beneficial than the inclusion of information from multiple linguistic domains. As such, training is likely to enhance perceptual sensitivities to the acoustic differences between tonal categories. As previously pointed out [2, 9], this may also alleviate the attentional and cognitive load associated with processing multi-domain linguistic information, especially for tone words which involve suprasegmental as well as segmental and lexical information. However, this does not rule out the contribution of semantic context in tone learning. Previous research has indicated that prosodic perception involves both bottom-up sensory-acoustic processing and higher-level linguistic processing, with the contribution varying as a function of linguistic experience [10]. It is thus conceivable that as learners gain more experience in perceiving subtle tonal distinctions, additional linguistic (contextual) information may be introduced in further training to enable learners to establish phonemic tonal categories that are generalizable at the word level.

5. NOTE

a. YW and AC contributed equally to this paper.

6. ACKNOWLEDGEMENTS

This research is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (312457-2006). We thank Caitlin Annable, Mathieu Dovan, Alison Kumpula, and Rebecca Simms for their assistance.

7. REFERENCES

[1] Davidson, L., Shaw, J., Adams, T. 2007. The effect of word learning on the perception of non-native consonant sequences. *JASA* 122, 3697-3709.

[2] Guion, S., Pederson, E. 2007. Investigating the role of attention in phonetic learning. In Bohn, O.-S., Munro, M. (eds.), *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins, 57-77.

[3] Hayes-Harb, R. 2007. Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23, 65-94.

[4] Hazan, V., Sennema, A., Iba, M., Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Comm.* 47, 360-378.

[5] Hirata, Y. 2004. Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *JASA* 116, 2384-2394.

[6] Logan, J., Lively, S., Pisoni, D.B. 1991. Training Japanese listeners to identify /r/ and /l/: A first report. *JASA*, 89, 874-886.

[7] Snodgrass, J.G., Vanderwart, M. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol.: Human* 6(2), 174-215.

[8] Wang, Y., Spence, M., Jongman, A., Sereno, J. 1999. Training American listeners to perceive Mandarin tones. *JASA* 106, 3649-3658.

[9] Werker, J.F., Fennell, C.T. 2004. Listening to sounds versus listening to words: Early steps in word learning. In Hall, G., Waxman, S. (eds.), *Weaving a lexicon*. MIT Press, 79-109.

[10] Zatorre, R.J., Gandour, J.T. 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society*, 1-18.