Distributional learning and overnight consolidation of a difficult non-native tone contrast

[1]Yin-To CHUI, [2]Zhen QIN

The Hong Kong University of Science and Technology
[1]ytchuiac@connect.ust.hk, [2]hmzqin@ust.hk

ABSTRACT

Previous research has demonstrated the success of distributional training for the acquisition of both segmental and suprasegmental categories, but no study has tested the distributional learning effect of a difficult tone contrast after overnight consolidation. The perception of Mandarin level tone (T1) – falling tone (T4) contrast has been shown to be especially difficult for native (L1) Cantonese speakers. The present study addresses whether this tonal contrast is successfully learned with distributional training, and whether this training effect is enhanced after overnight consolidation. 62 L1-Cantonese speakers were divided into two groups, where they were exposed to two different frequency distributions of the T1-T4 continuum (bimodal vs unimodal). Results showed slightly better post-training performance by the bimodal group, and this difference is enhanced after overnight sleep. This shows that the effect of overnight consolidation extends to statistical learning of lexical tones.

Keywords: Distributional training, overnight consolidation, lexical tone, Mandarin

1. INTRODUCTION

The distributional learning paradigm first appeared in [1] and followed a line of research on statistical learning mechanisms of language [2, 3]. Distributional learning refers to learning to discriminate between two phonetic categories (e.g. /d/ vs /t/) purely by implicit exposure to distributional information along a phonetic continuum (e.g. Voice onset time, VOT). As illustrated in Figure 1, a bimodal distribution consisting of more typical tokens of the two categories was found to benefit subsequent discrimination performance, as opposed to a unimodal distribution consisting of ambiguous tokens in between.

This ability to learn implicitly from frequency distributions has been shown to apply to infants [1, 4] and adults [5, 6], where it was demonstrated for consonants [1], vowels [4, 5], and lexical tones [6]. However, this seemingly ubiquitous nature of distributional learning is in fact not universal, as it has been demonstrated to be modulated by factors like learners' linguistic background. For example, the distributional learning of a Dutch vowel contrast /ɑ/-/a:/ might work for L1-Spanish speakers [5] but not L1-English speakers [7].

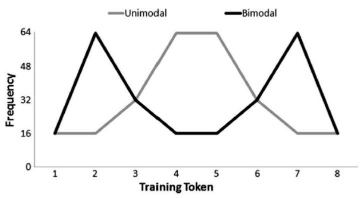

Figure 1: **Frequency counts of each token as heard in the unimodal and bimodal conditions [6]**

One aim of the present study is to see whether the distributional learning of tonal contrasts may also be subject to the influence of prior linguistic experience. Specifically, while previous research has demonstrated the success of distributional training of non-native tonal categories for non-tonal speakers (i.e. L1-English) [6, 8], the same might not be concluded for tonal speakers (e.g. L1-Cantonese), since the two populations exhibit differential processing difficulties of non-native tones. For example, the Mandarin level tone (T1) – falling tone (T4) contrast has been shown to be especially difficult for L1-Cantonese speakers, even more so than L1-English speakers [9]. Specifically, both tones in the contrast are perceptually assimilated to the same Cantonese high level tone, falling into the [Single Category] or [Category Goodness] classification according to the Perceptual Assimilation Model (PAM), which PAM posits to be hard to distinguish. This difficulty was verified empirically in a series of perception and production tasks, where L1-Cantonese speakers performed worse on this contrast than L1-English speakers [9].

The first goal of the present study, therefore, is to establish whether there is a distributional learning effect of the Mandarin T1–T4 contrast for L1 speakers of Cantonese. While previous studies have shown successful distributional learning of this contrast for L1-English speakers [8], it may turn out to fail for L1-Cantonese speakers given the added difficulty that arises from linguistic background.

Another unanswered question is whether the benefit of distributional training, if any, is enhanced after overnight consolidation. Memory consolidation refers to the stabilization of newly learned memory traces by transformation into long-term memory. Literature on sleep and learning suggests that sleep facilitates learning a new linguistic rule [10]; specifically, sleep that comes shortly after training facilitates the stabilization of newly-learned novel tonal categories, for example Cantonese tones by L1-Mandarin speakers [11]. Previous studies on distributional learning have focused on measuring only immediate performance after training. However, it is arguable that successful language learning should be indexed not only by immediate performance, but also whether the performance can be retained, so it is beneficial to identify whether consolidation benefits distributional learning as well where statistical learning mechanisms come into place. So far, only one distributional training study tested the longitudinal learning effect by retesting participants at 6- and 12-month intervals [12]. The results showed a lasting effect of distributional learning even in delayed intervals, but the influence of natural exposure outside the laboratory cannot be ruled out in that study.

The second goal of the present study, therefore, is to test whether the distributional learning effect (i.e. the performance of the bimodal over the unimodal group) is enhanced after overnight sleep. The benefit of sleep has been documented in other statistical learning paradigms [13], and the results of the current study can inform whether overnight consolidation extends to distributional learning.

2. METHODS

2.1. Participants

62 young adult native speakers of Cantonese were recruited in Hong Kong. Initial screening ensured that they had less than 9 years of Mandarin training and a self-reported proficiency of 'intermediate' or below. A cognitive battery was administered for matching purposes, which tested their pitch threshold [14], pitch memory [15], musical aptitude [16], working memory [17], and Mandarin vocabulary size [18]. A summary is given in Table 1.

	Unimodal	Bimodal
No. of participants	30 (9 males)	32 (15 males)
Age (year)	19.4 (1.37)	20.0 (1.49)
Pitch threshold test – Glide (in semitones)	4.6 (4.2)	4.3 (3.4)
Pitch memory task (in number of pitch notes)	5.4 (1.2)	5.2 (1.4)
MBEA (Musical aptitude)	0.83 (0.07)	0.85 (0.06)
Operation Span (Working memory)	60.9 (14.4)	63.2 (7.7)
PPVT(Mandarin-vocabulary size)	0.83 (0.07)	0.82 (0.08)

Table 1: Means (and standard deviations, in parentheses) of cognitive battery variables; MBEA: Montreal Battery of Evaluation of Amusia; PPVT: Peabody Picture Vocabulary Test

2.2. Stimuli

2.2.1. Discrimination Pre-test/post-test

A female and male native Mandarin speaker were invited to provide multiple recording tokens of two novel syllables /nua/ and /fao/ in Mandarin T1 and T4 (i.e. solid lines in Figure 2). Three tokens of each gender-syllable-tone combination were used in the discrimination task (see Procedure). The stimuli were novel to minimize the effect of prior exposure, and multiple syllables and genders were included to test participants' generalization of their discrimination performance to untrained settings (since only female /nua/ was used in the training phase).

2.2.2. Distributional training

The eight-step continuum used to construct the bimodal/unimodal distributions (Figure 1) was created using Praat. Only female /nua/ was used in training. Pitch contours of the naturally produced tones constituted the two end points of the continuum, and then 6 equidistant intermediate steps were synthesized in between using the ProsodyPro script in Praat [19], with the following procedure: 10 interpolation points were identified along the pitch contours (at 0%, 10%, …, 90%), then the distances between corresponding points were divided into seven equal spaces (Figure 2). The duration was normalized to the mean duration of the naturally produced tones, and the intensity was normalized to 65 dB SPL. Five native Mandarin speakers listened to the stimuli and confirmed that they are acceptable tokens of the syllable, and categorized the endpoints as good exemplars of T1 and T4. In addition to the syllable tokens, 32 sine-wave beeps were interspersed between the tokens for attended listening (see Procedure).

2.3. Procedure

The experiment followed a pre-test, training, and post-test procedure. Besides the T1-T4 pair, T1-T2 (i.e., a level-rising tonal contrast) was also tested for control purposes (stimuli created in the same way as described in section 2.2), and the results are not reported here given the space limit.

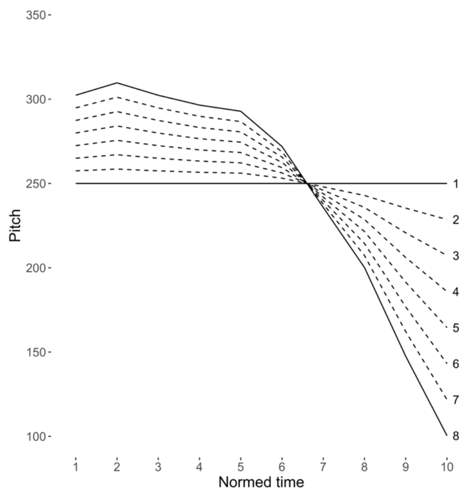

Figure 2: Mandarin T1-T4 continuum. Dashed line = intermediate tokens. Numbers = step number.

Participants first performed ABX tone discrimination tasks for the T1-T4 pair (order counter-balanced). In this task, subjects were exposed to natural T1 and T4 reference tokens A and B, followed by the target token X (taken from another recording) and was asked to identify whether X is the same tonal category as A or B in a 2-alternative forced choice (2AFC) task (Figure 3). All gender-syllable combinations were used here in a randomized, unblocked design. The inter-stimulus interval was chosen to be 1000ms, consistent with previous literature that this duration targets phonological categorization and processing. There were altogether 64 trials in each test. Participants' discrimination accuracy was chosen as the target measure.

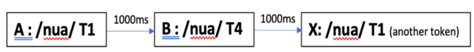

Figure 3: Example trial of the ABX discrimination task

Participants, matched in the cognitive tests, were then pseudo-randomly assigned to either the bimodal training group or the unimodal training group, where they were exposed to a bimodal tonal continuum or unimodal tonal continuum of /nua/ produced by the female talker. The syllable tokens were presented in a randomized order, with a total of 256 trials. In addition, 32 sine-wave beeps were interspersed in between the tokens, in which participants were instructed to press a keyboard response every time it appeared. This was to ensure the participants were actively attending to the stimuli, which has been shown to be necessary to induce distributional learning [6]. Interstimulus interval was set at 750 ms, and each training phase took around 10 minutes to complete.

Immediately after training, the discrimination task was assigned again (i.e. post-test 1). They were then instructed to wear a Fitbit device to record their sleep quality for the night, and they came back the next morning, after 12 hours, for the third round of ABX discrimination tasks (i.e. post-test 2).

2.4. Data analysis

Mixed-effects logistic regression models were performed on participants' response accuracy in the ABX discrimination tasks (binary-coded, 1 for correct and 0 for incorrect). Non-responded trials were excluded. The models were fitted in R using the lme4 package. Fixed effect is distribution (two levels: bimodal vs unimodal; deviation coding: -0.5, 0.5). To compare the performance of the bimodal and unimodal groups for each session (pre-test, post-test 1, and post-test 2), three nested models were performed with the effect of distribution nested within each session [20]. In all models, trial was entered as random effect.

3. RESULTS

Accuracy of the two groups in the ABX discrimination tasks is illustrated in Figure 4. The graph shows a steadily improving performance by the bimodal group across sessions, and a stable performance for the unimodal group.

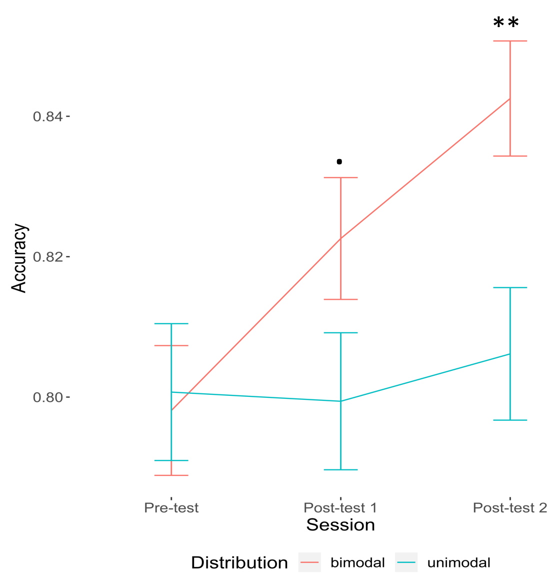

Figure 4: Accuracy (and standard errors) in ABX discrimination task by distribution and by session

A summary of the three nested models is presented in Table 2. The first nested model shows that the two groups' difference in accuracy in the pre-test did not reach significance, showing that the bimodal and unimodal groups had comparable baseline performance before distributional training. The second nested model shows that in post-test 1, immediately after training, the performance of the two groups started to diverge with marginal significance, with the bimodal group performing better than the unimodal group. This difference became highly significant in post-test 2 after overnight consolidation.

Session	Variable	Est.	SE	*z*	*p*
Pre-test	Distribution	0.016	0.084	0.195	0.845
Post-test 1	Distribution	-0.151	0.085	-1.78	**0.075**
Post-test 2	Distribution	-0.252	0.086	-2.916	**<0.01**

Table 2: Mixed effects logistic models on all participants' accuracy in the ABX discrimination task

4. DISCUSSION AND CONCLUSION

This study investigated whether L1-Cantonese speakers can benefit from distributional training of a Mandarin tone pair that has been shown to be difficult due to the influence of existing Cantonese tonal categories, and also whether any distributional learning effects can be enhanced after overnight consolidation. Overall results suggest a weak improvement after distributional training, and a stronger, enhanced improvement after overnight sleep.

In terms of immediate distributional learning performance, results only yielded a marginal improvement of the bimodal group over the unimodal group. When compared with the more robust success of distributional training exhibited in other adult studies testing different phonetic categories [5], or with a similar study testing the same Mandarin tonal contrast on another cohort (L1-English speakers without any prior tonal experience) [8], the weakened distributional learning performance by L1-Cantonese learners may indeed lend further support that distributional learning is modulated by prior linguistic experience, which, in the current case, might have arisen from the relative perceptual difficulties experienced by different L1 speakers as suggested by PAM. To make a stronger conclusion for this case, a direct comparison study of two participant groups of different linguistic backgrounds (i.e. L1-Cantonese vs L1-English) is needed in future research. It is also worth noting that we allowed our participants to have up to 9 years of Mandarin training, and amount of L2 experience may have played a role in the results, as participants' baseline performance may have been higher than naïve learners to begin with, making immediate improvement more difficult. More research is needed to confirm the effect of perceptual assimilation on the weakened effect of distributional learning for naïve learners, as it has been suggested that naïve learners will show an even stronger effect of perceptual assimilation [21] than experienced learners [9, 22], since naïve learners may rely purely on acoustic similarity, but experienced learners may pay more attention to the phonemic status of the target contrast and find it easier to distinguish between them [23].

In terms of performance after overnight consolidation, results showed that only the bimodal group continued to improve overnight, while the unimodal group stagnated at a stable performance level. This shows that looking at immediate performance may not enough to conclude the presence or absence of a distributional learning effect, as it is shown that even the weakened immediate performance is enhanced and reaches significance at a delay. The bottom line is that the distributional learning effect is there, just that it needed consolidation to make it visible. It will be interesting to see if the lack of effect in other studies can be recuperated if the learners are allowed to consolidate the newly-learned categories through an overnight interval. Current results also imply that overnight consolidation, while demonstrated in other more general training paradigms [11], may extend to the distributional learning of lexical tones.

Overall, the present findings suggest that Cantonese speakers exhibit a weak distributional learning effect of Mandarin tones, which might be due to influence by prior linguistic experience, and that overnight consolidation extends to the distributional learning setting and may be able to uncover distributional learning effects that were thought to be absent.

ACKNOWLEDGEMENTS

The project was funded in part by the Postgraduate Research Grant 2021-22 at the Hong Kong University of Science and Technology. We thank members of our research team, Ms. Ruofan Wu, Ms. Yuqi Wang for helpful comments and suggestions.

5. REFERENCES

[1] Maye, J., Werker, J. F., Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3), B101-B111.

[2] Saffran, J. R., Aslin, R. N., & Newport, E. L. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294), 1926-1928.
[3] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044), 606-608.
[4] Wanrooij, K., Boersma, P., & Van Zuijen, T. L. 2014. Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in psychology* 5, 77.
[5] Escudero, P., Benders, T., & Wanrooij, K. 2011. Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America* 130(4), EL206-EL212.
[6] Ong, J. H., Burnham, D., & Escudero, P. 2015. Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PloS one* 10(7), e0133446.
[7] Terry, J., Ong, J. H., & Escudero, P. 2015. Passive distributional learning of non-native vowel contrasts does not work for all listeners. *Proc. 17th ICPhS*, Glasgow, UK.
[8] Liu, L., Yuan, C., Ong, J. H., Tuninetti, A., Antoniou, M., Cutler, A., & Escudero, P. 2022. Learning to Perceive Non-Native Tones via Distributional Training: Effects of Task and Acoustic Cue Weighting. *Brain Sciences* 12(5), 559.
[9] Hao, Y. C. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics* 40(2), 269-279.
[10] Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. 2014. Sleep facilitates learning a new linguistic rule. *Neuropsychologia* 65, 169-179.
[11] Qin, Z., & Zhang, C. 2019. The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PloS one* 14(12), e0221498.
[12] Escudero, P., & Williams, D. 2014. Distributional learning has immediate and long-lasting effects. *Cognition* 133(2), 408-413.
[13] Simon, K. N., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L., & Gómez, R. L. 2017. Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain and Language* 167, 3-12.
[14] Qin, Z., Zhang, C., & Wang, W. S. Y. 2021. The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America* 149(1), 435-446.
[15] Williamson, V. J., & Stewart, L. 2010. Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory* 18(6), 657-669.
[16] Peretz, I., Champod, A. S., & Hyde, K. 2003. Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences* 999(1), 58-75.
[17] Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. 2005. An automated version of the operation span task. *Behavior research methods* 37(3), 498-505.
[18] Tsoi, E. Y. L., Yang, W., Chan, A., & Kidd, E. (2019). Mandarin–English speaking bilingual and Mandarin speaking monolingual children's comprehension of relative clauses. *Applied Psycholinguistics* 40(4), 933-964.
[19] Xu, Y. 2013. ProsodyPro—A tool for large-scale systematic prosody analysis. Laboratoire Parole et Langage, France.
[20] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. 2009. Mixed Effects Modelling for Nested Data. In: *Mixed effects models and extensions in ecology with R.* Springer New York, 101-142.
[21] So, C. K., & Best, C. T. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and speech* 53(2), 273-293.
[22] Leung, A. 2008. Tonal assimilation patterns of Cantonese L2 speakers of Mandarin in the perception and production of Mandarin tones. *Proc. 2008 CLA annual conference.*
[23] Wu, X., Munro, M. J., & Wang, Y. 2014. Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics* 46, 86-100.