Conditions on Adaptation to an Unfamiliar Lexical Tone System: The Role of Quantity and Quality of Exposure

Liang Zhao[1], Shayne Sloggett[1], and Eleanor Chodroff[1,2]

[1]University of York, Dept. of Language and Linguistic Science,

[2]University of Zurich, Dept. of Computational Linguistics

liang.zhao@york.ac.uk, shayne.sloggett@york.ac.uk, eleanor.chodroff@uzh.ch

ABSTRACT

Variation in lexical tone systems across dialects presents a perceptual challenge to speech adaptation. For instance, several Mandarin regional dialects have four phonological tones, but the phonetic realization of those tones differs considerably. Previous studies have demonstrated that listeners readily accommodate dialectal variation in lexical tone systems through incidental exposure. The present study examined whether adaptation was facilitated through the presence of explicit minimal-pair tone contrasts or increased exposure to the dialect. We found that rapid adaptation to the novel tone system was persistent even when minimal-pair sentences were removed from the stimuli and about one minute of incidental exposure was available with no repetition. Minimal-pair contrast was not necessary for adaptation. Increased exposure through repeated trials reliably enhanced sensitivity to the novel tone system.

Keywords: speech perception, lexical tone, perceptual adaptation, tone perception, Mandarin dialects

1. INTRODUCTION

Perceptual adaptation to unfamiliar speech requires adequate exposure to the target speech especially when the initial input is received later in adulthood. What makes the exposure "*adequate*" has become a central question to research on speech adaptation. Previous findings suggest that the adaptation outcome may be modulated by both quality and quantity of the spoken stimuli.

Quality refers to the type, structure and source of information in the experimental exposure. For example, lexical information has been considered beneficial for adaptation to a novel sound contrast [1, 2]. Hayes-Harb [2] tested English speakers' discrimination of /g/- and /k/-like novel sounds after auditory training with either minimal pairs of the sounds or with members of a [g]-[k] continuum without lexical meaning in a bimodal distribution that favored tokens towards the endpoints of the continuum. The results showed that adaptation to the novel contrast occurred with statistical learning alone, but discrimination was significantly enhanced when a lexical contrast was present. That said, listeners are able to rapidly discriminate between previously unheard linguistic contrasts: listeners are able to generalize heard patterns to new segments [3] and words unheard in the training [4]. Nevertheless, it is likely that the adaptation process could be facilitated if an explicit lexical contrast, such as a minimal pair, was heard in the exposure.

Listeners also frequently rely on explicit training to gather ample information for adaptation. If more information is available in the test stimuli, such as in sentences and passages, adaptation may happen without explicit training. Clarke & Garrett [5] reported listeners' rapid adaptation to Spanish- and Chinese-accented speech with one minute of incidental exposure to the sentence stimuli. In addition, exposure with about 16 sentences was found sufficient to initiate adaptation to a foreign-accented talker [6]. Crucially, however, listeners significantly improved over the course of the experiment. Even though adaptation can be reasonably successful in a short period, increased exposure may help listeners to generalize heard patterns to novel sounds or speakers.

In the present experiment, we investigated aspects of quality and quantity of incidental exposure in adaptation to a novel lexical tone system. Recently, Zhao, Sloggett, & Chodroff [7] identified that native Standard Mandarin listeners adapted to a novel lexical tone system from the Chengdu Mandarin dialect with less than two minutes of incidental exposure. Chengdu Mandarin has the same underlying four-tone system as Standard Mandarin, but disparate phonetic tone realizations (Figure 1) [8, 9, 10]. In the experiment, 24 sentence frames were manipulated to have either a semantically plausible (low-surprisal) or an implausible (high-surprisal) meaning by altering the tone category of one target word in the sentence-medial or sentence-final position. Native speakers of Standard Mandarin were asked to judge whether the spoken sentence was plausible or not with a "yes" or a "no" response in both their native dialect, Standard Mandarin, and the unfamiliar dialect, Chengdu Mandarin. Accurate judgment on the sentence plausibility and/or different response times between the high- and low-surprisal conditions indicated awareness of the surprisal tone manipulation and adaptation to the unfamiliar tone system. The results showed that listeners consistently slowed down for the high-surprisal sentences in Chengdu Mandarin starting from the beginning of the experiment, which strongly indicated rapid adaptation to the unfamiliar phonetic tones without explicit exposure to the dialect. This finding also leads to the hypothesis that phonetic tone information is always processed — even if such information may have little influence in lexical decision [11]. This has broader implications for models of speech perception involving lexical tone [11, 14, 15].

A critical aspect in the design of the experiment was that participants heard both the low- and high-surprisal versions of the sentence in each dialect, which provided minimal-pair sentences that contrasted in semantic plausibility and lexical tone category [7]. The presentation of minimal-pair sentences may facilitate rapid adaptation to a novel tone system. To test the potential conditions for adapting to a novel tone system with incidental exposure, the present study investigated 1) whether adaptation can still be achieved when minimal pairs are removed, and 2) if increasing the amount of incidental exposure would facilitate adaptation. We removed the minimal-pair contrast in the dialect-specific stimuli and introduced three repetitions of all trials in the new experiment. Minimal-pair presentation may be necessary for adaptation, in which case, we would expect that listeners have comparable response times for Chengdu Mandarin between high- and low-surprisal sentences, at least in the initial trials. However, a difference in response time may still emerge as incidental exposure increases with repetition.

The present study first investigated the effect of repetition on adaptation without minimal-pair presentation via the main experiment. To single out the effect of minimal-pair presentation, the response-time data in the first repetition block from the present experiment was extracted and compared to that from the previous experiment [7]. The absence of minimal-pair sentences was expected to impede adaptation. However, the effect of non-minimal-pair presentation was expected to be overcome by increasing the amount of ambient exposure through repetition.

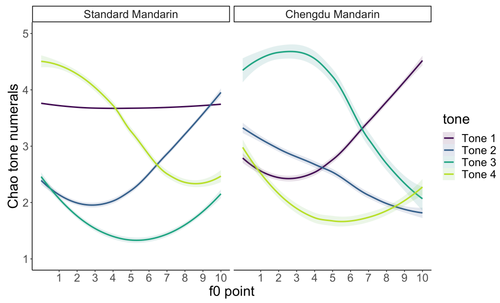

Figure 1: Smoothed lexical tone contours of Standard Mandarin and Chengdu Mandarin converted to Chao Tone numerals [10]. Ribbons reflect ±1 standard error of the mean.

2. METHODS

The current experiment replicated the design of the previous experiment of Standard Mandarin and Chengdu Mandarin [7], but with three repetition blocks and without the presence of minimal pair sentences in each dialect.

2.1. Participants

Thirteen native speakers of Standard Mandarin who reported little or no knowledge of Chengdu Mandarin participated in the experiment. No participant reported any hearing or reading impairments.

2.2. Materials

The experiment used the same 24 sentence frames as in Zhao et al. [7]. For each sentence frame, the lexical tone category of one target word was manipulated to have a semantically plausible (high-surprisal) or an implausible (low-surprisal) meaning (see Table 1). To avoid having both low- and high-surprisal versions of the sentence presented in the same dialect, each surprisal version was assigned to a different dialect. In addition, each critical-word tone category and position (medial or final, balanced evenly) were presented approximately the same number of times in each dialect.

2.3. Procedure

The experiment was built using Gorilla Experiment Builder [12]. Participants were asked to complete the semantic plausibility judgment task on their personal device with internet access in a quiet room, and with headphones if possible. The participants were made aware that they would be listening to sentences spoken in either the familiar Standard Mandarin or an unfamiliar Mandarin dialect. They were then presented with a test audio so they could adjust the volume of the sound output to a comfortable level.

The practice phase consisted of two trials that introduced the test phase procedure. To start the trial, the participant pressed a "play audio" button on the screen. After listening, participants responded to the question *"Does this sentence make sense"* by clicking on either the "yes" or "no" response button on the screen. Feedback was then provided by displaying the orthographic form of the sentence on the screen. The two trials contained one high-surprisal and one low-surprisal Standard Mandarin sentence. These sentences were not repeated in the main experiment.

In the test phase, the participants received one of the two lists described in the materials. The task was identical to the practice trials except that no feedback was given. There were 24 trials in each block, and a total of three blocks for each dialect (24 trials × 3 repetitions × 2 dialects). The participants heard Chengdu Mandarin blocks first, then the Standard Mandarin blocks to avoid task-based learning effects in the adaptation process due to prior familiarity with the sentence frames in the native dialect. The trials in each repetition block were randomized.

low-surprisal sentence	a) 有 一只 鹰 在 天上 飞 *You3 yi4 zhi1 ying1 zai4 tian1 shang4 fei1* *There is an eagle in the sky flying* "There is an eagle flying in the sky"
high-surprisal sentence	b)* 有 一只 鹰 在 天上 肥* *You3 yi4 zhi1 ying1 zai4 tian1 shang4 fei2** *There is an eagle in the sky gaining weight** "There is an eagle gaining weight in the sky"

Table 1: An example sentence item across surprisal conditions.

2.4. Data analysis

The effects of *surprisal* (high surprisal vs. low surprisal), *dialect* (Standard Mandarin vs. Chengdu Mandarin) and *repetition* were assessed on accuracy and response time. Responses that matched the expected plausibility judgment were considered *correct*: "yes" responses to low-surprisal plausible sentences and "no" responses to high-surprisal implausible sentences. Response time was calculated as the interval between the end of the audio file and the click registering a judgment.

Accuracy was modeled with a Bayesian logistic mixed-effects regression, and response time with a Bayesian log-normal mixed-effects regression, both with weakly informative priors [13]. Each model included fixed effects of surprisal, dialect, two repetition contrasts, and the full set of interactions. The random effect structure for participants included an intercept and slopes for surprisal, dialect, repetition contrasts and all the interactions, and for sentence frame, an intercept and a random slope for surprisal. The priors for the accuracy model were $N(0, 20)$ for the intercept, main effects and interactions, and $N(0, 0.05)$ for random effects. For the response-time model, the priors were $N(7, 1)$ for the intercept, $N(0, 1)$ for main effects and interactions, and $N(0, 0.01)$ for random effects. The model was run for 2000 iterations with a burn-in period of 1000 iterations. Surprisal and dialect were sum-coded (*surprisal:* high-surprisal $= 1$, low-surprisal $= -1$; *dialect:*

Chengdu Mandarin = 1, Standard Mandarin = −1). Reverse Helmert coding was used for the three-level repetition factor, comparing one level to the mean of the previous level(s) (*repetition contrast 1*: block 2 = 1/2, block 1= -1/2, block 3 = 0; *repetition contrast 2*: block 3 = 2/3, block 2 = − 1/3, block 1 = − 1/3). An estimate was deemed credible in its direction of influence on the dependent variable if the 95% credible interval excluded 0 (i.e., no effect).

For the comparison between the present and previous experiments, response time was modeled with fixed effects of surprisal, dialect, trial, presentation (*presentation*: with-minimal-pair design = 1, no-minimal-pair design = −1), and the interaction of surprisal, dialect and presentation. The random effects for participants included an intercept and slopes for surprisal and dialect, and for frame an intercept and a slope for surprisal. The same priors were used as the above response-time model.

3. RESULTS

3.1. Results from the current experiment

3.1.2. Accuracy

The overall accuracy across dialect and surprisal conditions (Figure 2) closely resembled the previous findings: 1) high near-ceiling accuracy in both surprisal conditions for the familiar dialect and 2) considerably lower accuracy in the high-surprisal condition for the unfamiliar dialect, Chengdu Mandarin. A gradual improvement can be seen over the three blocks in the Chengdu high-surprisal conditions.

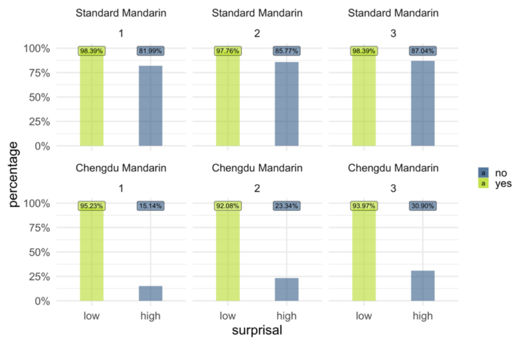

Figure 2: Percentage of *correct* responses across dialect, surprisal and repetition (“1, 2, 3” refer to the repetition blocks).

The model revealed credible main effects of surprisal, dialect, and the interaction between dialect and surprisal as in the previous study. Specifically, accuracy was higher in the low-surprisal condition than in the high-surprisal condition (*surprisal:* β = −7.64, 95% CI = [−14.89, −3.10]); accuracy was higher for sentences spoken in Standard Mandarin than in Chengdu Mandarin (*dialect:* β = −7.04, 95% CI = [−14.24, −2.51]); the credible interaction between surprisal and dialect indicated higher accuracy for Standard Mandarin low-surprisal sentences (*dialect x surprisal:* β = 5.13, 95% CI = [0.58, 12.32]). For the effect of repetition, accuracy was reliably different from the first to the second block of repetition (*repetition contrast 1*: β = −7.78, 95% CI = [−21.09, −0.19]), but not from the first two blocks to the third block (*repetition contrast 2*: β = 4.79, 95% CI = [−6.06, 19.63]). The second repetition block reliably interacted with surprisal (*surprisal x repetition contrast 1*: β = 8.29, 95% CI = [0.73, 21.59]) and dialect (*dialect x repetition contrast 1*: β = 7.70, 95% CI = [0.11, 20.95]), while the third repetition block showed no credible interaction with the other factors (*surprisal x repetition contrast 2*: β = −4.23, 95% CI = [−18.96, 6.62]; *dialect x repetition contrast 2*: β = −4.39, 95% CI = [−19.14, 6.43]; *surprisal x dialect x repetition contrast 2*: β = 4.63, 95% CI = [−6.13, 19.48]). This suggested that accuracy improved after the second repetition of the trials for the high-surprisal sentences compared to the low-surprisal, and for Chengdu sentences compared to Standard Mandarin, but these did not reliably improve in the third block.

3.1.2. Response time

The response-time model identified credible effects of all tested factors and their interactions, except for the interaction between surprisal and the second repetition contrast. To be exact, the credible effects of surprisal (*surprisal:* β = 0.15, 95% CI = [0.12, 0.18]), dialect (*dialect:* β = 0.11, 95% CI = [0.08, 0.13]) and the interaction between surprisal and dialect (*surprisal x dialect:* β = −0.07, 95% CI = [−0.09, −0.05]) replicated the patterns found in the previous study: listeners were reliably slower in the high-surprisal condition in both dialects, with greater difference between the surprisal conditions in Standard Mandarin than in Chengdu Mandarin (Figure 3). Nevertheless, a difference was still observed between high- and low-surprisal conditions in Chengdu Mandarin.

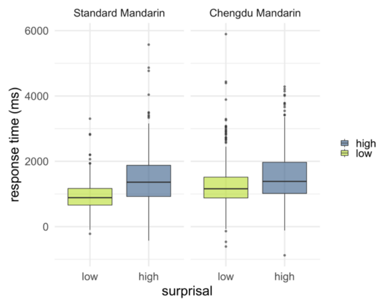

Figure 3: Response times across dialect and surprisal conditions.

For the effect of repetition (Figure 4), all responses generally accelerated block by block (*repetition contrast 1*: β = −0.16, 95% CI = [−0.21, −0.11]; *repetition contrast 2*: β = −0.19, 95% CI = [−0.24, −0.14]). Moreover, slower responses were found for Chengdu sentences after each repetition, relative to Standard Mandarin sentences (*dialect x repetition contrast 1*: β = 0.11, 95% CI = [0.06, 0.17]; *dialect x repetition contrast 2*: β = 0.10, 95% CI = [0.05, 0.14]). However, response times were credibly faster for high-surprisal sentences from the first to the second block (*surprisal x repetition contrast 1*: β = −0.06, 95% CI = [−0.11, −0.003]), possibly driven by the faster responses to Standard Mandarin sentences, but there was no difference towards the third block (*surprisal x repetition contrast 2*: β = −0.02, 95% CI = [−0.07, 0.02]). In fact, response times were reliably modulated by the three-way interactions between surprisal, dialect and both repetition contrasts (*surprisal x dialect x repetition contrast 1*: β = 0.07, 95% CI = [0.02, 0.13]; *surprisal x dialect x repetition contrast 2*: β = 0.06, 95% CI = [0.01, 0.11]), suggesting block-wise slowdown for Chengdu high-surprisal sentences, but block-wise speed-up for Standard Mandarin high-surprisal sentences.

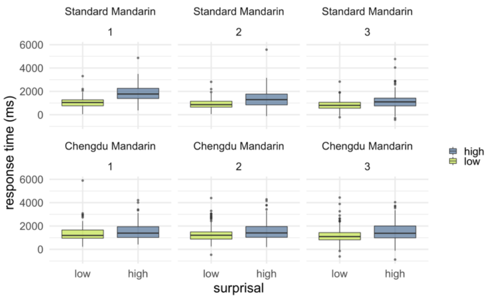

Figure 4: Response times across dialect, surprisal and repetition conditions ("1, 2, 3" refer to the repetition blocks).

3.2. The effect of minimal pairs

To examine the effect of minimal-pair presentation on response time, we compared the first block of data in the present study to the data in the previous study, which only differed in the presence of minimal pairs (Figure 5). Accuracy was not examined as surprisal did not influence responses to Chengdu sentences in either study. No credible effect of presentation was detected between the two designs (*presentation*: β = 0.05, 95% CI = [−0.03, 0.12]), indicating that the effect of minimal-pair presentation was not as salient as expected. Response times were consistently slower for high-surprisal sentences in both experiments (*surprisal*: β = 0.20, 95% CI = [0.15, 0.25]), reflecting listeners' awareness of surprisal manipulation even without minimal pairs or repetition. The high-surprisal slowdown did not reliably interact with presentation (*surprisal x presentation*: β = 0.01, 95% CI = [−0.03, 0.06]; *surprisal x dialect x presentation*: β = −0.0006, 95% CI = [−0.04, 0.03]), indicating that the removal of minimal pairs did not credibly affect listeners' sensitivity to the surprisal manipulation. Nevertheless, the estimated mean difference between high and low surprisal differed numerically between the two sets of data in the expected direction: about 170 ms without minimal pairs, and 270 ms with minimal pairs. Minimal-pair presentation may have numerically facilitated adaptation to the novel tone system, resulting in greater distinction between the surprisal manipulations; removal of the minimal pairs reduced, but did not obviate the effect of surprisal. There was also a credible interaction between presentation and dialect (*dialect x presentation*: β = 0.06, 95% CI = [0.003, 0.11]), indicating slower responses to Chengdu sentences when minimal pairs were present.

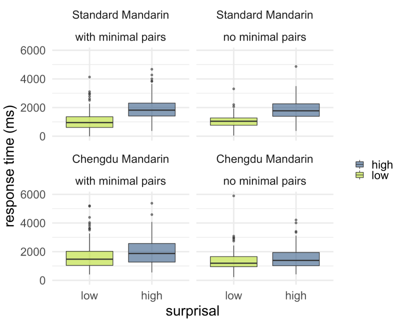

Figure 5: Response times across dialect, surprisal and presentation conditions in the previous (with-minimal-pair) and the new (no-minimal-pair) experiments

4. DISCUSSION

The previous experiment on Chengdu Mandarin revealed rapid adaptation to the novel tone system with incidental exposure containing minimal-pair sentences [7]. The current experiment found that rapid adaptation to the novel tone system was persistent even when minimal-pair sentences were removed from the stimuli and only minimal incidental exposure was available. Enhancement in quality (minimal pairs) and quantity (repetition) of the exposure both facilitated adaptation in terms of accuracy and response time, but they were not necessary factors.

The present study showed similar results for the effects of dialect and surprisal as in the previous study, which indicated successful adaptation and shared perceptual mechanisms for novel tone processing with [7] and without minimal pairs. Increased incidental exposure reliably boosted adaptation to the unfamiliar tone system, evidenced by improved accuracy and increased difference in response times between the low- and high-surprisal conditions starting from the second block of repetition.

Specifically for the effect of minimal-pair presentation, it was surprising to find that listeners were sensitive to the surprisal manipulations even when the minimal pairs were removed from exposure. In fact, adaptation occurred under rather adverse conditions, where incidental exposure was limited to one repetition and with no minimal-pair sentences in the same dialect. Nevertheless, inclusion of minimal pairs in the stimuli can assist discrimination between the low- and high-surprisal meanings, and may potentially direct more attention to the tone contrast and ease the process of adaptation or learning of the new tone system.

As minimal incidental exposure was sufficient for adapting to the unfamiliar tone system with or without minimal pairs in the exposure, it is unlikely that listeners relied on increased exposure to initiate adaptation. Repetition was more likely a consolidating factor as the mappings between the phonological categories and the novel phonetic tone realizations were reinforced over repetitions.

5. CONCLUSION

The present study investigated the potential factors impacting adaptation to an unfamiliar tone system by testing the effects of minimal-pair presentation of the stimuli and increased incidental exposure. We found that both factors were adequate to induce adaptation, though neither was necessary. The current study also has implications of more readily processed tones for models of speech perception. Further analysis should delve into the tone-specific adaptation which was not reported due to limited space. As the current experiment examined adaptation with incidental exposure, further research could be done to assess the effect of explicit training for adaptation to unfamiliar lexical tone systems.

6. ACKNOWLEDGMENTS

The final author was supported by SNSF grant 208460. The authors would also like to thank Sam Hellmuth and Paul Foulkes for helpful feedback.

7. REFERENCES

[1] Norris, D., McQueen, J. M., & Cutler, A. "Perceptual learning in speech," *Cognitive Psychology*, *47*(2), 204–238. 2003.

[2] Hayes-Harb, R. "Lexical and statistical evidence in the acquisition of second language phonemes," *Second Language Research*, 23(1), 65-94. 2007.

[3] Maye, J., Aslin, R. N., & Tanenhaus, M. K. "The weckud wetch of the wast: Lexical adaptation to a novel accent," *Cognitive Science*, *32*(3), 543–562. 2008.

[4] McQueen, J. M., Cutler, A., & Norris, D. "Phonological abstraction in the mental lexicon," *Cognitive science*, 30(6), 1113-1126. 2006.

[5] Clarke, C. M., & Garrett, M. F. "Rapid adaptation to foreign-accented English," *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. 2004.

[6] Bradlow, A. R., & Bent, T. "Perceptual adaptation to non-native speech," *Cognition*, *106*(2), 707–729. 2008.

[7] Zhao, L., Sloggett, S., & Chodroff, E. "Top-down and Bottom-up Processing of Familiar and Unfamiliar Mandarin Dialect Tone Systems," in *Proc. Speech Prosody*, May.2022, pp. 842-846.

[8] Hou, J.Y. (侯精一). *The Modern Outline of Chinese Dialects* (现代汉语方言概论). China: Shanghai Education Press (上海教育出版社), 2002.

[9] Li, Rong. (李荣). "Chengdu Dialect Dictionary (成都方言词典)," *The Modern Dictionaries of Chinese Dialects* (现代汉语方言大词典). China: Jiangsu Education Press (江苏教育出版社), 2002.

[10] Zhao, L., & Chodroff, E. "The ManDi Corpus: A Spoken Corpus of Mandarin Regional Dialects," in *Proc. 13th Language Resources and Evaluation Conference*, May.2022, pp. 1985-1990.

[11] Gao, X., Yan, T. T., Tang, D. L., Huang, T., Shu, H., Nan, Y., & Zhang, Y. X. "What makes lexical tone special: a reverse accessing model for tonal speech perception," *Frontiers in psychology*, 2830 (10). 2019.

[12] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. "Gorilla in our midst: An online behavioral experiment builder," *Behavior Research Methods*, *52*(1), 388–407. 2020.

[13] Bürkner, P.-C. (n.d.). *Advanced Bayesian Multilevel Modeling with the R Package brms*. Aug.2022. Available: https://github.com/paul-buerkner/brms.

[14] Strauss, T. J., Harris, H. D., & Magnuson, J. S. "jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition," *Behavior Research Methods*, 39(1), 19-30. 2007.

[15] L. Shuai and J. G. Malins, "Encoding lexical tones in jTRACE: a simulation of monosyllabic spoken word recognition in Mandarin Chinese," *Behavior Research Methods*, 49 (1), 230– 241. 2017.