Development of Category Perception between [n] and [l] in Standard Chinese among Southwest Mandarin Speakers

Jie Hou[1, 2], Yu Chen[1], and Jianwu Dang[2]

[1]Tianjin University of Technology

[2]Japan Advanced Institute of Science and Technology

December 25, 2023

Abstract

Southwestern Mandarin speakers often encounter challenges in distinguishing syllable-initial contrast between [n] and [l] in Standard Chinese. The present study aimed to explore the formation of non-native phonemes by examining the development of category perception in adult learners. Both Southwest Mandarin speakers and Standard Chinese speakers were recruited, and Southwest Mandarin speakers were trained on prevocalic [n] and [l] in a 50-day training program. The typical category perception experiments, including 2AFC identification and AX discrimination, were conducted to assess the participants' perceptual patterns and progress. Results indicated that speech training facilitated the maturation of perceptual patterns among adult learners. Significant changes were observed in both identification and discrimination, with greater improvements in identification than discrimination. The findings suggest that non-native speech learning among adult learners shares similarities with native speech development, supporting the Speech Learning Model (SLM) that learning L2 phonemes involves mechanisms and processes employed in acquiring the L1 sound system.

Keywords: Category Perception Development, Phoneme Confusion, Speech Training

1 INTRODUCTION

In the field of speech perception, Category Perception (CP) is a fundamental concept that explains how humans process speech by nonlinearly mapping continuous sound signals onto discrete phoneme segments (Liberman et al., 1957). Consequently, CP has greatly contributed to the exploration of physiological and psychological mechanisms of speech perception (Liberman & Mattingly, 1985) and the construction of computational models of speech perception

(Kleinschmidt & Jaeger, 2015; Pan et al., 2022). As categorical perception plays a critical role in first language (L1) acquisition, numerous studies have focused on the perception of various segments (consonants: Liberman et al. 1957; Larkey et al. 1978; vowels: Fry et al. 1962; Repp et al. 1979) and suprasegments (lexical tones: Xu et al. 2006; Peng et al. 2010), as well as the perceptual development of native speakers (Medina et al., 2010; Ma et al., 2021; Feng & Peng, 2023). Similarly, the mechanism and process of non-native speech perception, especially how second language (L2) learners develop the ability to split one L1 phoneme into two separate L2 phonemes, is also the hot spot that has received considerable attention in recent years.

The early perceptual development in infants and children contributes significantly to category learning. Many studies have reported a critical period of speech category learning between 2 and 12 months of age (Werker & Hensch, 2015). During this period, infants exhibit equal sensitivity to both native and non-native phonemes, enabling them to distinguish various speech sound contrasts with ease (Kuhl, 2004; Werker, 2018). Then, with growing exposure to a specific language, infants' capacity to discriminate native sounds is enhanced, while their ability to discriminate non-native sounds is weakened (Werker & Tees, 1984; Kuhl et al., 2006; Maurer & Werker, 2014). For instance, Kuhl et al. (2006) found that American and Japanese infants in 6-8 months showed similar discriminating responses to [r] and [l] in English. However, in 10-12 months, American infants showed improved performance, whereas Japanese infants showed reduced performance.

The canonical view posits that infants already form categories in their early development. However, an alternative perspective based on distributional learning algorithm with machine learning technology, has been proposed by Feldman et al. (2021) and Schatz et al. (2021). According to this view, infants do not learn 'speech categories', but instead learn 'auditory spaces'.

In other words, they collapse the phonetic-independent dimension so that the distribution of the phonetic-dependent dimension corresponds to the category space. This explanation is credible as infants younger than 12 months lack sufficient knowledge and experience of phonetic categories to assist perception. Subsequently, McMurray et al. (2018) demonstrated that categorization in children is still gradually developing from ages 7 to 18 thanks to an increasing understanding of speech variability. Similarly, in another study comparing the perception of Mandarin tones among four- to seven-year-old children, Chen et al. (2017) found that older children exhibited more adult-like identification (sharper slope) and discrimination (higher peak) functions. Given the immature state of infant speech development, category learning in humans will continue into childhood and adolescence.

Compared to infants and children, adults often experience considerable difficulty in distinguishing non-native phonemes. This difficulty has been attributed to factors such as low sensory plasticity or interference from their native phonemic inventories. According to the Critical Period Hypothesis (CPH), the relationship between learners' age and their susceptibility to L2 input is non-linear (Vanhove, 2013; Stölten et al., 2015). However, the Speech Learning Model (SLM) believes that the ability to establish phonetic categories remains intact across the lifespan (Flege, 1995). The SLM argues that learning difficulties are primarily due to the interaction of the learner's L1 and L2 phonological systems. Consequently, when L2 contrasts do not exist in the native phonetic system, adults may struggle to perceive or produce them accurately. For example, the perception of English [r] and [l] by Japanese speakers (Goto, 1971), the perception of English vowels by Spanish speakers (Flege et al., 1997), and the perception of Mandarin tones by non-tone language speakers (Hallé et al., 2004).

Although challenging, many previous studies have attempted to help adults access categories from non-native speech. These studies have primarily focused on how L2 learner build up L2 phoneme categories through language exposure (Miyawaki et al., 1975; MacKain et al., 1981), category training (synthetic stimuli, by manipulating different acoustic cues, Strange & Dittmann 1984; Iverson et al. 2005), and listening training (natural stimuli, by showing variability from different speakers or contexts, Shinohara & Iverson 2018; K. Zhang et al. 2018). Among these studies, one of the most famous topics is how Japanese speakers learn to separate [r] and [l] into two phonemes in English. MacKain et al. (1981) found that Japanese speakers with extensive experience in English conversation performed significantly better than those without such experience, though their discrimination performance was still inferior to that of native English speakers. By providing same-different discrimination training with immediate feedback, Strange & Dittmann (1984) found that learners could achieve gradual but effective improvements in identification and discrimination after 14 to 18 training sessions. Overall, these studies raise the possibility of perceptual development in adults through substantial training.

Extensive research on category perception of non-native speakers has yielded valuable insights into various training techniques and their effectiveness (Ingvalson et al., 2014; Sakai & Moorman, 2018; Nagle & Baese-Berk, 2022). However, most studies have been carried out within a short training schedule, typically ranging from 7 to 14 sessions. Moreover, previous research has also suggested that short-term exposure is insufficient to induce dramatic changes in the categorization of non-native phonemes. Consequently, little attention has been given to the step-by-step perceptual development of adult learners to track the formation of non-native speech categories.

With that in mind, the current study determined to explore Chinese dialect speakers' development of categorical perception of [n] and [l] in Standard Chinese with a 50-day period of speech training. Significantly, this study takes the contrast between [n] and [l] in Standard Chinese as the research object. There are highly similar acoustic properties and articulatory characteristics between [n] and [l]. As the sonorant consonants, both of them exhibit attributes that lie between vowels and consonants, and can function as flexible components within one syllable (onset, nucleus, or coda) in many languages. Therefore, exploring the perception of these two consonants will firstly enhance our understanding of diverse speech segments. Furthermore, the voiced alveolar nasal consonant [n], and voiced alveolar lateral approximant [l], are phonetically and phonologically distinct in syllable-initial position in Standard Chinese (Yuan, 2001). However, many native speakers of Jianghuai Mandarin (Cao, 2008; Li et al., 2012), Southwest Mandarin (Cao, 2008; W. Zhang & Levis, 2021), and Northwest Mandarin (Yuan, 2001) cannot differentiate between the two consonants in either perception or production. For local people of these dialects, the confusion of [n] and [l] causes no misunderstanding in closed communication within their own community. However, these people would face prevailing difficulties when learning Standard Chinese or conducting cross-dialect communication. Some individuals may even be frustrated in job hunting or career development due to their poor speech intelligibility in Standard Chinese. Therefore, exploring the perceptual mechanism and training of [n] and [l] in Standard Chinese is of paramount importance to help people affected by phoneme confusion or special accent.

For this purpose, we recruited participants with both consistent or inconsistent distinction of [n] and [l]: 60 native speakers of Standard Chinese as the criterion group for categorization, and 20 native speakers of Southwestern Mandarin as the training group with an additional 50-day computer-based speech training. Then, the typical category perception experiments, including

2AFC identification and AX discrimination, will be conducted to measure the perceptual patterns and progress of participants. Finally, one set of data from native speakers of Standard Chinese and four sets of data from native speakers of Southwest Mandarin are obtained for further analysis and comparison. Time series with five nodes can be constructed by considering the criterion group as a delayed post-test (ideal expectation or forecast) of the training group.

More specifically, this work aims to address the following two research questions:

RQ1: What is the developmental process of adult category acquisition? The Speech Learning Model (SLM) of L2 speech acquisition proposes that the mechanisms and processes used in learning the L1 sound system would be applied to L2 learning. Thus, we assume that category acquisition in adults (non-native speech) would share certain commonalities with perceptual development from infancy to adolescence (native speech). Specifically, it is predicted that after 50 days of long-term training, adult learners will develop basic category perception between [n] and [l]. With the deepening of speech training, we expect to observe changes in both identification and discrimination (e.g., sharper slope, higher peak), as well as asynchronization between identification and discrimination.

RQ2: What are the possible factors that impede adult category acquisition? Despite numerous cross-sectional and longitudinal studies have documented the consistent or inconsistent distinction between [n] and [l] in China, limited attention has been given to the underlying causes of this phenomenon. This study would like to further discuss the specific factors that contribute to perceptual confusion and learning challenges associated with [n] and [l] in Standard Chinese. To achieve this, we will draw upon recent theories of L2 speech perception and learning, such as the Perceptual Assimilation Model (PAM) and the Second Language Linguistic Perception (L2LP) model, along with relevant research on acoustic and linguistic parameters.

2 METHOD

2.1 Participants

Twenty native speakers of Southwestern Mandarin from Hubei University were recruited for this study as the experimental group (10 females and 10 males, mean age ± SD: 20.7±1.59, age range: 18-24). They reported confusions of /n/-/l/ contrast in perception and production on a quick listening and speaking test.

The control group consisted of seventy students from Tianjin University of Technology (35 females and 35 males, mean age ± SD: 19.54±0.79, age range: 18-22). They all grew up in Northern China and speak Standard Chinese (Northern Mandarin) as their first language. All of them could perceive and produce [n] and [l] properly.

Overall, no participant reported any speech, language, or hearing diseases and disorders. All of them had signed the informed consent prior to the experiment and were paid for their participation.

2.2 Stimuli

An eleven-step speech continuum between Chinese syllable [le] and [ne] was generated by TANDEM STRAIGHT (Kawahara & Morise, 2011). The endpoints of the speech continuum were a standard syllable [le] and a standard syllable [ne] with high-level tone uttered by a male native speaker of Northern Mandarin. The endpoint sounds were recorded in wav format: mono soundtrack, 44100Hz sampling rate, 32-bit resolution and were aligned to have the same duration (500ms) and average intensity (72db).

Since there was no clear evidence to provide reliable acoustic cues responsible for distinguishing the two phonemes, the present study employed holistic morphing approach to gradually change acoustic information in both temporal and spectral. All fundamental frequency, formants (including frequency, amplitude, and bandwidth) and anti-formant structure were reserved (to keep natural and articulate) and altered in this process.

2.3 Procedure

2.3.1 Training and testing

The members of experimental group participated in a 50-day long-term speech training using real-time feedback technology (Guo et al., 2018). In production training, subjects were guided to pronounce different syllables beginning with [n] or [l] and will receive real-time feedback through smiling face or crying face. The feedbacks came from vibration frequency detection via a sensor attached to the subject's ala of nose. In perceptual training, subjects were guided to listen to and identify various syllables beginning with [n] or [l], then they will receive real-time feedback through smiling face or crying face.

The experimental group participated in the category perception experiment mentioned below four times during the training period, with an interval of about two weeks: one pre-test (day-0), two mid-tests (day-17 and day-34) and one post-test (day-50). Correspondingly, the control group participated in the experiment once to obtain the standard category perception model.

2.3.2 Identification task

Two-alternative forced choice (2AFC) paradigm was adopted in the identification task. Stimuli were presented singly to participants. Participants were asked to judge the initial consonants of the syllables they heard as [n] or [l] and press the corresponding bottom. The entire task consisted of a total of 110 randomly presented trials (11 stimuli × 10 repetitions). The formal task was preceded by a short practice block for participants to understand the experimental instructions. The practice block employed a different speech continuum, with recordings from a female native speaker of Northern Mandarin. The response mappings were counterbalanced across participants ('F' for [l], and 'J' for [n], or reversed).

2.3.3 Discrimination task

The discrimination task adopted AX paradigm with two-step difference and stimuli were presented in pairs with a 500-ms interstimulus interval (ISI). There were nine contrastive units in the speech continuum (1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11) and each unit had four possible combinations of the two elements (AA, BB, AB, BA). Participants were instructed to judge whether the initials of the two stimuli were the same or different and press the corresponding key. Each unit was repeated five times, resulting in 145 trials in total being presented to participant in random order (9 pairs × 4 combinations × 5 repetitions − 35 overlaps). The practice block and button counterbalancing ('F' for [same], and 'J' for [different], or reversed) were also applied in the discrimination task.

2.4 Data analysis

Statistical analysis was performed in R 4.2.0 (R Core Team, 2022). In the following statistics, we fitted generalized linear model using glm() function in stats package (R Core Team, 2022), linear mixed-effects model using lmer() function and generalized linear mixed model using glmer() function in lme4 package (Bates et al., 2015) and lmerTest package (Kuznetsova et al., 2017), and generalized additive model using bam() function in mgcv package (Wood, 2017). Akaike Information Criterion (AIC) was used in model comparison to ensure that all models mentioned above were optimal models. The model with mixed-effects were be chosen if random factor showed improvement to the model. Post Hoc comparisons or pairwise comparisons were conducted using emmeans package (Lenth, 2023), with degrees-of-freedom derived using Kenward-Roger method and p-values adjusted using Tukey method.

2.4.1 Identification

Identification Function A generalized linear model was fitted to the entire identification data, with response (1 = [l], 0 = [n]) as dependent variable, step number (numerical predictor, 1-11) and session (numerical predictor, from 0 to 4, here we set control group as 4) as independent variable, logit as link function. The same model without session factor was also applied to each participant under each session to examine individual differences in categorical perception across participants and sessions.

Position and Sharpness Two key parameters of category boundary were extracted from each fitted identification function: position and sharpness. The following equation was used to calculate the parameters (Xu et al., 2006):

$$\log_e\left(\frac{P_I}{1-P_I}\right) = b_0 + b_1 x \tag{1}$$

where P_I is identification score and x is step number. The sharpness is expressed as coefficient b_1, which refers to the slope of the transition part. When setting $P_I = 0.5$, the position x_{cb} can be calculated as $-b_0/b_1$, corresponding to 50% crossover point on identification function (See Eq. 2).

$$b_0 + b_1 x_{cb} = \log_e \left(\frac{0.5}{1 - 0.5} \right) \quad \Rightarrow \quad x_{cb} = -\frac{b_0}{b_1} \tag{2}$$

To further evaluate the progress on identification parameters, we adopted linear mixed-effects models, in which dependent variables were position and sharpness, independent variable was session, and random factor was by-subject intercepts. [formula: position or sharpness session + (1 | subject)].

2.4.2 Discrimination

Discrimination Function Considering the non-linear relationship between discrimination scores and stim units, the generalized additive mixed model was fitted to discrimination data to get discrimination curves, in which parametric term was session (linear predictor), smooth term was stim unit (non-linear predictor), and this model also included random smooths (equivalent to random slopes and intercepts) for participants.

Between-category and Within-category Discrimination To explore the relationship between between-category and within-category discrimination, we conducted a generalized linear mixed model, with response (1 = ‘different’, 0 = ‘same’) as dependent variable, stim type (categorical predictor, sum coding: -1 = ‘between category’ and 1 = ‘within category’) and session (numerical predictor, from 0 to 4, here we set control group as 4) as independent variable, and by-subject intercepts as random factor. For the stim type, the between-category included one comparison unit (5-7) corresponding to the categorical boundary; within-category included two comparison units (1-3 and 9-11) at the endpoints of the continuum.

2.4.3 Individual difference

In addition, two measurements were defined to assess each subject's progress in identification tasks and discrimination tasks:

Identification Accuracy: taking control group as criteria, we set correct response of step 1 to 5 as [l], correct response of step 7 to 11 as [n] and excluded response of step 6. The value was proportion of correct responses of the remaining 100 trials.

Discrimination Peak: the maximum response value on the discrimination curve and was accepted only if this value occurs in step units 4-6, 5-7 and 6-8.

3 RESULTS

3.1 Identification

The identification functions for each session were presented in Figure 2A. The position and sharpness of category boundary were shown in Figure 2B and Figure 2C respectively.

3.1.1 Identification functions

The response function of the matched control group displayed a typical category perception pattern characterized by a sigmoid shape with a steep slope. In contrast, the experimental group exhibited a weak downward linear trend before training (S0) and gradually approached the control group's identification curves as the training session progressed. Eventually, the response curve of the experimental group also approximated to the sigmoid shape, albeit with a gentler slope.

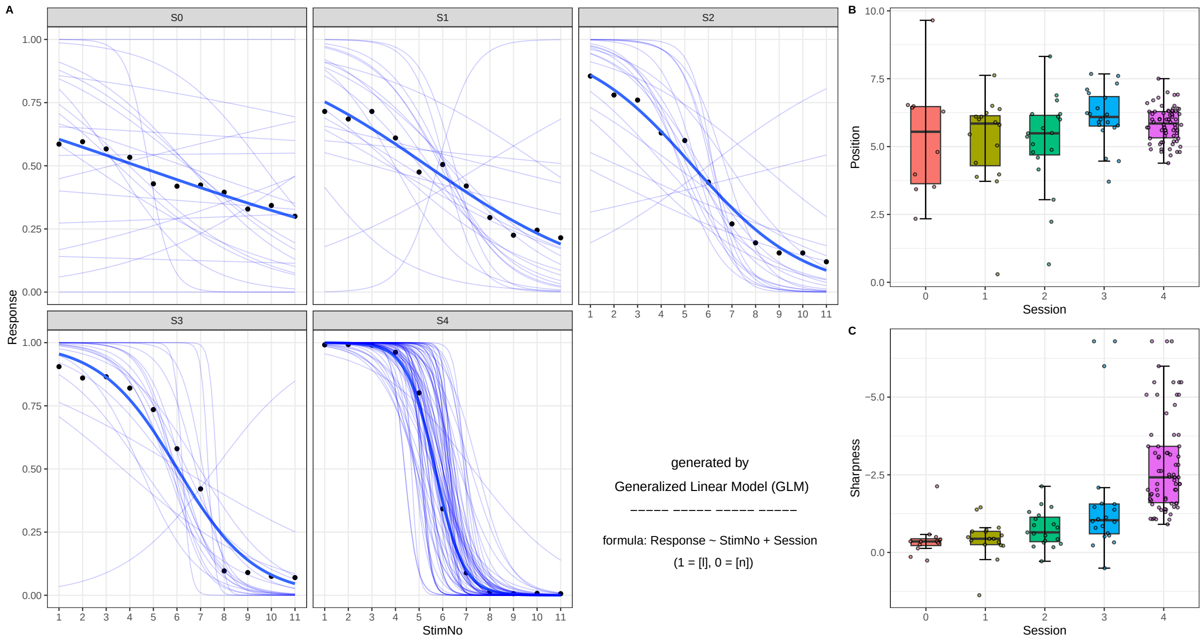

Figure 1: Identification functions and category boundary parameters. (A) the identification scores (black dots) and the fitted identification functions (blue lines) across training sessions; (B) position of category boundary; (C) sharpness of category boundary.

The generalized linear model revealed significant main effect of step number (β = -0.041, *SE* = 0.011, *z* = -3.613, *p* < .001), session (β = 1.629, *SE* = 0.038, *z* = 42.418, *p* < .001), as well as two-way interactions between step number and session (β = -0.279, *SE* = 0.006, *z* = -45.124, *p* < .001). These findings suggested that session was a strong predictor of the response across identification functions, thereby indicating a meaningful influence of speech training on participants' identification performance.

3.1.2 Position of category boundary

The mean boundary positions of all sessions fell between 5 and 7, which roughly corresponded to the acoustic boundary (6) on the speech continuum. In the linear mixed-effects model, no significant main effect of session was observed on position (β = 0.016, *SE* = 0.094, *t* = 0.167, *p* = 0.868). Although pairwise comparisons did not show any significant position differences across all sessions, we could still see that the post-training boundary positions were much closer to the values of the control group and had less variations (Pre: 5.56±2.04; Post: 6.09±1.00; Control: 5.86±0.67).

3.1.3 Sharpness of category boundary

The linear mixed-effects model showed a significant main effect of session on sharpness (β = -0.628, *SE* = 0.095, *t* = -6.599, *p* < .001). Furthermore, pairwise comparisons showed significant differences in sharpness between S0 and S3 (β = 1.203, *SE* = 0.404, *t* = 2.973, *p* < .05), and between S3 and S4 (β = 1.095, *SE* = 0.375, *t* = 2.922, *p* < .05). These findings suggested that the sharpness of category boundary in the experimental group greatly improved over the course of training. However, despite this improvement, the slope was still not as sharp as that of the control group.

3.2 Discrimination

3.2.1 Discrimination functions

Figure 3A showed the discrimination curves across all sessions. As we can see, discrimination accuracy in the control group was at chance-level in two respective categories but significantly increased near the category boundary. This same pattern could also be found in the training group, except for S0 and S1, where the response only slightly changed across all stimuli pairs. Namely, untrained participants had neither the highest peak at the center boundary, nor the lowest valleys at the two endpoints.

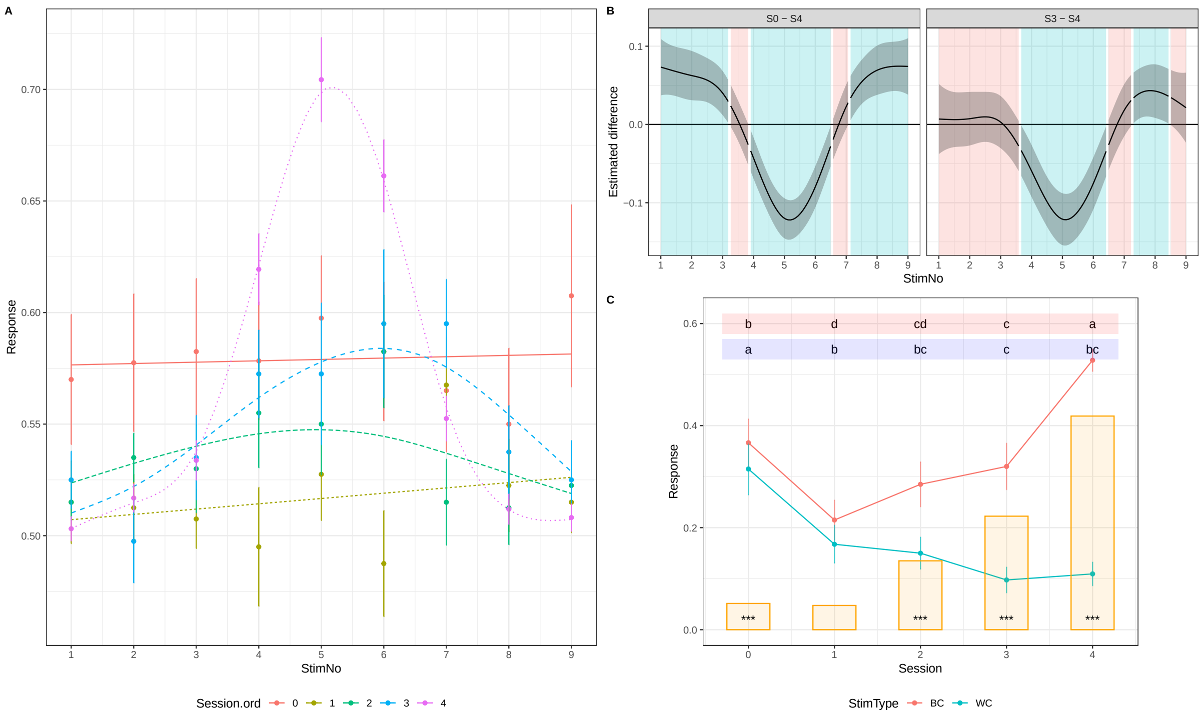

Figure 2: Discrimination functions and related parameters. (A) the fitted discrimination curves across training sessions; (B) left panel: the model estimates of difference between S0 and S4 (pre-test vs. control), right panel: the model estimates of difference between S3 and S4 (post-test vs control), the cyan shade represents statistical significance and the red shade represents statistical insignificance; (C) the response variation for different stimulus types: between-category (red line), within-category (cyan line) and difference between between-category and within-category (yellow bars).

The GAMM model based on straight line showed that discrimination response varied as a function of stimulus units, in which the S0 and S1 were linear (S0: EDF = 1.001, Ref.DF = 1.001, F = 0.041, p = 0.841; S1: EDF = 1.000, Ref.DF = 1.000, F = 0.621, p = 0.431), while the others were non-linear (S2: EDF = 1.877, Ref.DF = 2.338, F = 1.023, p = 0.354; S3: EDF = 2.756, Ref.DF = 3.421, F = 3.606, $p < .05$; S4: EDF = 6.712, Ref.DF = 7.555, F = 42.820, $p < .001$). The effective degrees of freedom (EDF) indicated non-linear changes (EDF > 1) in the discrimination trajectory.

Crucially, both control (S4: EDF = 6.712, Ref.DF = 7.555, F = 42.692, $p < .001$) and post-trained (S3: EDF = 2.756, Ref.DF = 3.421, F = 3.217, $p < .05$) trajectories differed from the pre-training trajectory (S0). The estimated difference between S0 and S4 revealed that the two trajectories differed from each other across almost the entire range (1.00 - 3.26, 3.83 - 6.56, 7.06 - 9.00), whereas the difference between S3 and S4 was mainly in the central part (3.59 - 6.50, 7.30 - 8.52).

3.2.2 Between- and within-category discrimination

The generalized linear mixed-effects model revealed significant main effect of stim type (β = 0.102, SE = 0.040, z = 2.519, $p < .05$), session (β = -0.300, SE = 0.029, z = -10.176, $p < .001$) and interactions between stim type and session (β = 0.269, SE = 0.016, z = 16.365, $p < .001$). The variations in stim type indicated that response sensitivities differed between between category and within category, and the changes in session indicated that response accuracies developed over time.

Simple effect results showed that, for between-category discrimination, S3 was significantly higher than S1 (β = 0.578, SE = 0.167, z = 3.454, $p < .01$), while still lower than controls (β =

-0.982, *SE* = 0.219, z = -4.492, $p < .001$). For within-category discrimination, S3 was significantly lower than S0 (β = -1.521, *SE* = 0.188, z = -8.099, $p < .001$), but no longer significantly different from controls (β = -0.086, *SE* = 0.263, z = -0.327, p = 0.9975). Additionally, the difference between between-category and within-category discrimination increased continuously throughout the training period and all showed strong statistical significance, except S1.

3.3 Individual and task differences

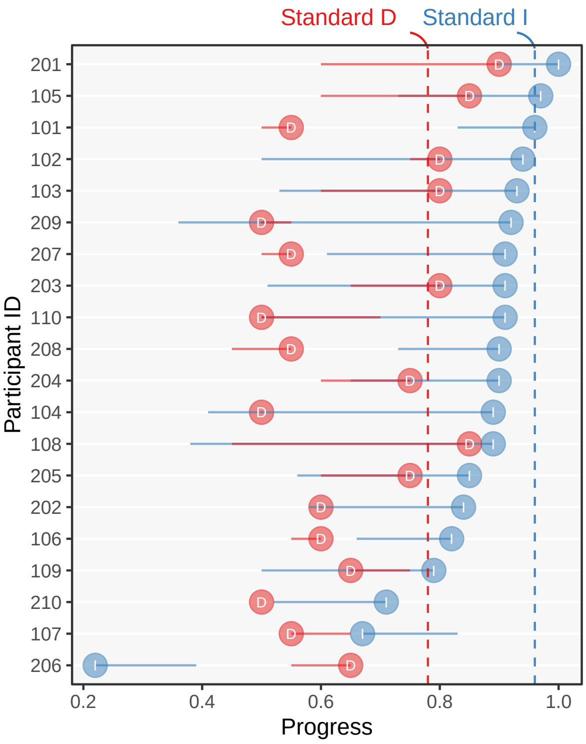

Figure 3: Individual changes in identification and discrimination. The left endpoint of the line depicted on the lollipop plot represents the value before training, whereas the right dot represents the value after training; the vertical red line indicates the mean identification accuracy of the standard group, while the vertical cyan line indicates the mean discrimination peak of the standard group.

Overall, all trained participants had made significant improvement in identification (S0: 58.4%, S1: 68.0%, S2: 77.3%, S3: 84.7%) and discrimination (S0: 75.0%, S1: 58.3%, S2: 62.8%, S3: 66.0%, note that S0 required special consideration), and the improvement in identification was greater than that in discrimination. For reference, the control group achieved high scores in both identification (96.0%) and discrimination (77.9%).

Figure 4 showed the individual performance in terms of identification and discrimination.

Identification: four participants (S201 to S102) successfully entered the standard interval; nine participants (S103 to S108) were very close to meeting the criterion value; then, five participants (S205 to S210) were still in the developmental phase, while two participants (S107 and S206) failed in the identification task.

Discrimination: eight participants performed equally or better than the norm; five participants made some progress but did not meet the target; conversely, three participants (S104, S202, S210) did not demonstrate any progress, and four participants (S209, S110, S109, S107) experienced retrogression.

A comparison of the two tasks revealed that most of the subjects developed asynchronously between identification and discrimination, with identification progressing earlier and faster than discrimination. Only one example was found where discrimination exceeded identification.

4 DISCUSSION

5 CONCLUSION

6 ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health [grant numbers xxxx]. We would like to express our great appreciation to Longchuan Li, Xiangjun Guo and Sen Yang for their previous efforts in this long-term research.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

10

Cao, Z. (2008). *Linguistic Atlas of Chinese Dialects [Hanyu Fangyan Dituji]*. Beijing: The Commercial Press.

6

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, *44*(6), 1413–1434. (Edition: 2016/12/05) doi: 10.1017/S0305000916000581

4

Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021). Do Infants Really Learn

Phonetic Categories? *Open Mind*, *5*, 113–131. Retrieved 2023-03-15, from https://doi.org/10.1162/opmi_a_00046 doi: 10.1162/opmi_a_00046

3

Feng, Y., & Peng, G. (2023). Development of categorical speech perception in Mandarin-speaking children and adolescents. *Child Development*, *94*(1), 28–43. Retrieved from https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/cdev.13837 doi: 10.1111/cdev.13837

3

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (Vol. 92, pp. 233–277). York Press.

4

Flege, J. E., Bohn, O.-S., & Jang, S. (1997, October). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095447097900528 doi: 10.1006/jpho.1997.0052

4

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech*, *5*(4), 171–189. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/002383096200500401 doi: 10.1177/002383096200500401

3

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds ”l” and ”r.”. *Neuropsychologia*, *9*, 317–323. doi: 10.1016/0028-3932(71)90027-3

4

Guo, X., Chen, Y., Dang, J., Li, L., & Qiu, Q. (2018). A pilot study of phonetic category learning considering speech production mechanisms: data of mandarin’s [l] and [n] by dialect speakers. In *Studies on speech production: 11th international seminar, issp 2017, tianjin, china.*

9

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004, July). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095447003000160 doi: 10.1016/S0095-4470(03)00016-0

4

Ingvalson, E. M., Ettlinger, M., & Wong, P. C. M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, *18*(1), 35–47. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1367006912456586 doi: 10.1177/1367006912456586

5

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. doi: 10.1121/1.2062307

5

Kawahara, H., & Morise, M. (2011). Technical foundations of tandem-straight, a speech analysis, modification and synthesis framework. *Sadhana - Academy Proceedings in Engineering Sciences*, *36*. doi: 10.1007/s12046-011-0043-3

8

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi: 10.1037/a0038695

3

Kuhl, P. K. (2004, November). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. Retrieved from https://doi.org/10.1038/nrn1533 doi: 10.1038/nrn1533

3

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7687.2006.00468.x doi: 10.1111/j.1467-7687.2006.00468.x

3

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

10

Larkey, L. S., Wald, J., & Strange, W. (1978, July). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, *23*(4), 299–312. Retrieved from https://doi.org/10.3758/BF03199713 doi: 10.3758/BF03199713

3

Lenth, R. V. (2023). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.8.5)

11

Li, B., Zhang, C., & Wayland, R. (2012, November). Acoustic Characteristics and Distribution of Variants of /l/ in the Nanjing Dialect*. *Journal of Quantitative Linguistics*, *19*(4), 281–300. Retrieved from https://doi.org/10.1080/09296174.2012.714537 doi: 10.1080/09296174.2012.714537

6

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368. doi: 10.1037/h0044417

2, 3

Liberman, A. M., & Mattingly, I. G. (1985, October). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. Retrieved from https://www.sciencedirect.com/science/article/pii/0010027785900216 doi: 10.1016/0010-0277(85)90021-6

2

Ma, J., Zhu, J., Yang, Y., & Chen, F. (2021, July). The Development of Categorical Perception of

Segments and Suprasegments in Mandarin-Speaking Preschoolers. *Frontiers in Psychology*, *12*. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.693366 doi: 10.3389/fpsyg.2021.693366

3

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390. (Edition: 2008/11/28) doi: 10.1017/S0142716400009796

5

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.21177 doi: 10.1002/dev.21177

3

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*, 1472–1491. doi: 10.1037/dev0000542

4

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010, October). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, *38*(4), 493–503. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095447010000446 doi: 10.1016/j.wocn.2010.06.002

3

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O.

(1975, September). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340. Retrieved from https://doi.org/10.3758/BF03211209 doi: 10.3758/BF03211209

5

Nagle, C. L., & Baese-Berk, M. M. (2022). ADVANCING THE STATE OF THE ART IN L2 SPEECH PERCEPTION-PRODUCTION RESEARCH: REVISITING THEORETICAL ASSUMPTIONS AND METHODOLOGICAL PRACTICES. *Studies in Second Language Acquisition*, *44*(2), 580–605. (Edition: 2021/07/16) doi: 10.1017/S0272263121000371

5

Pan, L., Ke, H., & Styles, S. J. (2022, March). Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *Scientific Reports*, *12*(1), 4703. Retrieved from https://doi.org/10.1038/s41598-022-08557-7 doi: 10.1038/s41598-022-08557-7

3

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S. Y. (2010, October). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095447010000707 doi: 10.1016/j.wocn.2010.09.003

3

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

10

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 129–145. doi: 10.1037/0096-1523.5.1.129

3

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224. (Edition: 2017/10/30) doi: 10.1017/S0142716417000418

5

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, *118*(7), e2001844118. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.2001844118 doi: 10.1073/pnas.2001844118

3

Shinohara, Y., & Iverson, P. (2018, January). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242–251. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095447017300530 doi: 10.1016/j.wocn.2017.11.002

5

Strange, W., & Dittmann, S. (1984, March). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145. Retrieved from https://doi.org/10.3758/BF03202673 doi: 10.3758/BF03202673

5

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). EFFECTS OF AGE AND SPEAKING RATE ON VOICE ONSET TIME: The Production of Voiceless Stops by Near-Native L2 Speakers. *Studies in Second Language Acquisition*, *37*(1), 71–100. (Edition: 2014/06/02) doi: 10.1017/S0272263114000151

4

Vanhove, J. (2013). The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis. *PLOS ONE*, *8*(7), e69172. Retrieved from https://doi.org/10.1371/journal.pone.0069172 doi: 10.1371/journal.pone.0069172

4

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*(4), 703–728. (Edition: 2018/09/11) doi: 10.1017/S0142716418000152

3

Werker, J. F., & Hensch, T. K. (2015). Critical Periods in Speech Perception: New Directions. *Annual Review of Psychology*, *66*(1), 173–196. Retrieved from https://www.annualreviews.org/doi/abs/10.1146/annurev-psych-010814-015104 doi: 10.1146/annurev-psych-010814-015104

3

Werker, J. F., & Tees, R. C. (1984, January). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and De-*

velopment, *7*(1), 49–63. Retrieved from https://www.sciencedirect.com/science/article/pii/S0163638384800223 doi: 10.1016/S0163-6383(84)80022-3

3

Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman and Hall/CRC.

11

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. Retrieved from https://asa.scitation.org/doi/abs/10.1121/1.2213572 doi: 10.1121/1.2213572

3, 11

Yuan, J. (2001). *Outline of Chinese dialects [Hanyu Fangyan Gaiyao]*. Beijing: Language and Culture Press.

6

Zhang, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S.-Y. (2018, October). The Effect of Speech Variability on Tonal Language Speakers' Second Language Lexical Tone Learning. *Frontiers in Psychology*, *9*. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01982 doi: 10.3389/fpsyg.2018.01982

5

Zhang, W., & Levis, J. M. (2021, August). The Southwestern Mandarin /n/-/l/ Merger: Effects on Production in Standard Mandarin and English. *Frontiers in Communication*, *6*. Retrieved from https://www.frontiersin.org/articles/10.3389/fcomm.2021.639390 doi: 10.3389/fcomm.2021.639390

6