EVIDENCE FOR PRESERVED PHONETIC SENSITIVITY TO LEXICAL TONES IN CHINESE INTERNATIONAL ADOPTEES

Andrea C.-L. Chang[1], Tyler K. Perrachione[1], & Sung-Joo Lim[1,2]

[1]Department of Speech, Language, and Hearing Sciences, Boston University, USA
[2]Department of Psychology, Binghamton University, USA
sungjoo@binghamton.edu

ABSTRACT

During infancy, children gain perceptual sensitivity to phonetic contrasts in their native language while losing sensitivity to foreign contrasts. It remains unclear whether traces of language exposure in infancy remain after sudden and permanent interruption of heritage language acquisition, as in international adoption. We examined sensitivity to lexical tones in adults (age 18-23 years) with a history of international adoption (age 3-18 months) whose heritage language was Chinese. We assessed phonetic, phonological, and lexical sensitivity to Chinese tones using tone contour identification and discrimination, auditory Stroop, and Chinese sentence identification tasks, and the auditory brainstem frequency-following response (FFR). International adoptees had native-like FFR, and their tone discrimination and identification were intermediate between English and Mandarin control groups. However, higher-level processing of Chinese did not differ between adoptees and the English control group. Exposure to heritage languages in infancy may create traces of early phonetic environments that persist into adulthood.

Keywords: tonal language perception, speech perception, critical period, frequency-following response

1. INTRODUCTION

Infants become sensitive to the speech sounds of their native language within the first year of life. They exhibit enhanced sensitivities to native speech contrasts [4] while losing sensitivity to foreign-language speech contrasts [11]. Such “neural commitment” [3] to one’s native-language speech contrasts suggests that early linguistic experience may have a significant impact on lifelong tuning of speech processing.

However, it remains unclear whether traces of language exposure from early infancy remain even after sudden and radical changes in the linguistic environment, as in international adoption. The existing evidence is equivocal as to whether early language experiences leave traces that persist across development [7,10]. Nevertheless, two recent EEG studies demonstrated that children adopted from China retain neural sensitivity to Mandarin lexical tones [6,9]. Based on this evidence from cortical recordings, it is possible that the impact of early language exposure to Mandarin also leaves persistent traces of neural sensitivity to the phonetic representations of lexical tones at the subcortical level (i.e., auditory brainstem) [2].

To this end, the present study examined phonetic, phonological, and lexical sensitivity to Mandarin Chinese lexical tones using a battery of behavioral tests in normal hearing adults who were adopted from China compared to native English and native Chinese controls. We also examined differences in the sensitivity of the subcortical auditory system in these groups via FFRs elicited by Mandarin lexical tones.

2. METHODS

2.1. Participants

We recruited three participant groups with different language backgrounds: English-speaking adults who had been adopted from China to the U.S. as children (n=9; mean age = 20 years), native Mandarin speakers (n=21; mean age = 22.2), and native American English speakers with no tone language experience (n=21; mean age = 21.2). On average, the adoptees arrived in the U.S. at 9.6 months of age (Table 1). Adoptees’ heritage language cannot be known for certain and may have been Mandarin, Cantonese, or another dialect; six of the adoptees had some post-adoption exposure to Mandarin (usually in college). We quantified their Mandarin knowledge using the *Hanyu Shuiping Kaoshi* (HSK) Level 1 test [1].

All participants had normal hearing and provided written informed consent, approved by the Institutional Review Board at Boston University.

Subj.	Age (yr.)	Adoption age (mo.)	Mandarin Experience (yr.)	HSK score
P1	19	6	0	N/A
P2	21	3	0	N/A
P3	18	10	1	28/40
P4	21	6	6	32/40
P5	19	12	3	23/40
P6	20	12	11	39/40
P7	20	18	2	39/40
P8	23	8	2	14/40
P9	21	11	0	N/A

Table 1: International adoptees’ demographics.

2.2. Behavioral tasks

Participants performed all behavioral tasks in a sound attenuated booth. All auditory stimuli were RMS normalized to 70 dB SPL.

2.2.1. Pitch-contour perception test (PCPT)

Participants performed a non-lexical pitch pattern identification task, which is designed to assess listeners' aptitude for learning lexical tones [8,12]. Five vowels (/a/, /i/, /o/, /e/, and /u/) recorded by four native American English speakers (2 females) were resynthesized to have three Mandarin tonal contours (level, rising, and falling) using the pitch synchronous overlap-and-add algorithm (PSOLA) [5]. On each trial, listeners heard one vowel and matched its synthesized pitch contour to a corresponding arrow icon using assigned keys on a response box. Participants completed 120 trials (4 speakers × 5 vowels × 2 repetitions). No feedback was provided.

2.2.2. Tone discrimination task

Participants performed an AX lexical tone discrimination task on pairs of Mandarin monosyllables and synthesized sinewave tones. On each trial, participants heard a pair of stimuli that differed in consonant, vowel, and speaker, and they judged whether the pitch contours of the pair were the same (e.g., mǎi/yě) or different (e.g., tīng/xià). The speech stimuli were recorded by one female and one male native Mandarin speaker. The sinewave tones were synthesized from the natural speech sounds using PSOLA.

Participants completed 192 trials (half with same and half with different tone contours) for both speech and non-speech stimuli. No feedback was provided.

2.2.3. Auditory Stroop task

To assess international adoptees' implicit semantic processing of their heritage language (whether as a result of latent knowledge or due to re-exposure in adulthood), we conducted an auditory Stroop task. On each trial, listeners heard one of the two words meaning "high" or "low" spoken in English, Mandarin, and Cantonese; each word was spoken in either an exaggeratedly high- or low-pitched voice by one male native speaker of each language. Participants judged whether the voice pitch was high or low on each trial, and we hypothesized that participants' familiarity with a language would be reflected in slower response times (semantic interference) when a word's meaning and its voice pitch were incongruent vs. congruent.

Participants performed 40 trials (20 congruent) per language condition. Corrective feedback was provided on each trial.

2.2.4. Sentence identification task

We conducted a sentence identification task to assess whether the adoptees retained enhanced familiarity for recognizing their heritage language (or had gained such familiarity from post-adoption exposure). In this task, participants heard sentences spoken in Mandarin, Cantonese, Japanese, Korean, or Arabic, and were asked to identify whether the sentence was "Chinese" or not. The sentences were taken from "The North Wind and Sun" passage recorded by four native speakers of each language (2 females, 2 males). Listeners performed a total of 80 trials (4 sentences × 4 speakers × 5 languages) presented in a random order. No feedback was provided.

2.3. Frequency-following response (FFR)

Four minimally contrastive Mandarin monosyllables (/yi1/ "clothing", /yi2/ "aunt", /yi3/ "chair", /yi4/ "easy"), recorded by one native Mandarin speaker, were used to elicit the FFR. In an electrically shielded, sound attenuated booth, participants passively listened to randomly presented syllables while watching a silent movie. The stimuli were delivered via Etymotic ER-1 insert earphones at 70 dB SPL.

A total of 6400 trials were collected from each participant (1600 trials per stimulus). All stimuli were presented at alternating polarities.

2.3.1. Data acquisition, preprocessing, and analyses

We measured FFR using EEG, recorded from a standard 10/20 layout of 32 active scalp electrodes (Biosemi ActiveTwo) at a sampling rate of 4096 Hz. Two reference electrodes were placed on the earlobes. Electrode offsets for all channels were maintained below 30 μV. EEG data were high-pass filtered at 70 Hz cutoff to isolate subcortical responses from cortical activity. Trials were removed if the activity range exceeded 35 μV. The remaining trials for each stimulus per participant (M ± SD = 1475 ± 133 trials) were averaged for further analyses.

For each participant, we extracted f_0 contours for the stimuli and FFR via autocorrelation using a 40-ms sliding window (3-ms steps). We quantified the pitch tracking accuracy of the FFR by computing stimulus-to-response correlation; this measure indexed how well the FFR tracked the f_0 of the stimulus.

3. RESULTS

3.1. Pitch-contour perception test

We used a generalized linear mixed-effects model to examine whether participants' language background had an impact on identifying pitch contours of vowels

(Fig. 1). Mandarin controls performed significantly better on the PCPT than did international adoptees (z = 2.85, p = 0.004); and the adoptees had significantly higher accuracy than the native English listeners (z = 2.00, p = 0.046).

Next, we examined whether the pitch identification accuracy of the adoptees depended on their age of adoption or Mandarin proficiency (HSK score). These correlations did not reveal any significant relationships (both $p > 0.57$).

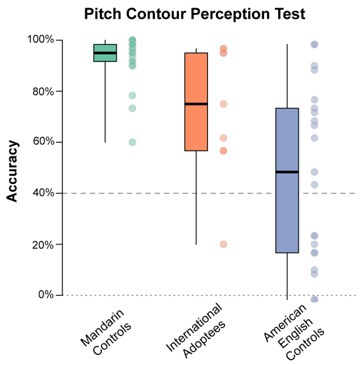

Figure 1: Pitch contour identification accuracy by participants in each group.

3.2. Tone discrimination performance

We used linear mixed-effects models to examine whether the participants in each group differed in their ability to discriminating spoken Mandarin lexical tones and their sinewave analogs (Fig. 2).

This analysis revealed significant main effects of group ($\chi^2(2)$ = 89.35, p < 0.001) and stimulus type ($\chi^2(2)$ = 18.25, p < 0.001), as well as a significant interaction between group and stimulus type ($\chi^2(2)$ = 84.99, p < 0.001). The linear model with pairwise contrast coding revealed that Mandarin controls performed significantly better than the international adoptees overall (z = 6.24, p < 0.001). Mandarin controls also had a significant advantage for discriminating speech vs. sinewave tones compared to the international adoptees (z = 5.77, p < 0.001). Furthermore, the adoptees performed significantly better than native English group overall (z = 4.05, p < 0.001), and likewise had a relative advantage for discriminating speech vs. sinewave tones (z = 2.06, p = 0.040).

We also examined whether the adoptees' tone discrimination was related to their age of adoption or HSK score. While HSK scores were positively correlated with speech tone discrimination (r = 0.78, p = 0.01), this relationship was not found for sinewave tones (p = 0.72). Additionally, age of adoption was marginally correlated with speech tone discrimination (r = 0.63, p = 0.068), but was not related to discrimination of sinewave tones (p = 0.42).

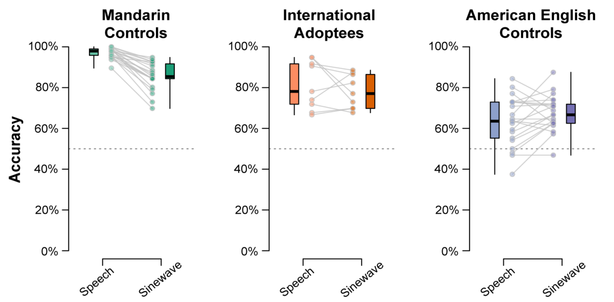

Figure 2: Accuracy discriminating Mandarin speech and sinewave tones, by group.

3.3. Auditory Stroop task performance

Table 2 shows the amount of interference by group and language condition when the word meaning and vocal pitch were incongruent (i.e., "high" said with a low-pitched voice) compared to when they were congruent (i.e., "high" said with a high-pitched voice).

	Mandarin Controls	International Adoptees	Native English
Mandarin	55 ±110	31 ± 72	16 ± 73
Cantonese	67 ± 103	–7 ± 46	–13 ± 70
English	79 ± 176	54 ± 73	102 ± 79

Table 2: Response time (RT) differences (mean ± s.d. in ms) in the incongruent vs. congruent trials in each language condition and participant group.

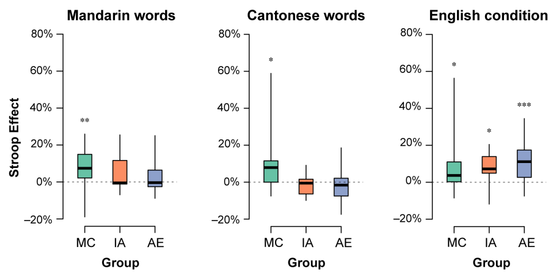

Figure 3: Semantic interference in each language condition by group. Auditory Stroop effect: (mean incongruent RT – mean congruent RT) ÷ mean congruent RT.

We performed a separate linear mixed-effects model on RT for incongruent vs. congruent trials for each language condition and each participant group (Fig. 3). When identifying the voice pitch of Mandarin words, Mandarin controls were significantly slower to respond to incongruent vs. congruent trials (t_{772} = 3.36, p < 0.001), but no significant interference (Stroop effect) was found in the other two groups (both $p \geq 0.23$). Likewise for the Cantonese words, only Mandarin controls exhibited a significant Stroop effect (t_{771} = 3.30, p = 0.001). In the English condition, all groups exhibited significant Stroop effects (Mandarin controls: t_{748} = 2.42, p = 0.02; Adoptees: t_{331} = 2.56, p = 0.01, native English: t_{777} = 5.13, p < 0.0001).

The magnitude of international adoptees' Stroop effect for Mandarin words was not correlated with their age of adoption or HSK score (both $p > 0.17$).

3.4. Sentence identification performances

As expected, Mandarin controls were highly accurate at identifying Chinese sentences (i.e., Mandarin and Cantonese) vs. other languages (Fig. 4). The linear mixed-model analysis revealed that Mandarin controls exhibited significantly higher rates of identifying Mandarin and Cantonese sentences as Chinese compared to both the adoptee ($z = 2.07$, $p = 0.038$) and native English groups ($z = 3.67$, $p < 0.001$). In contrast, the international adoptees and native English groups misidentified Korean and Japanese sentences as Chinese significantly more often than did Mandarin controls (both $p < 0.001$).

We found that correct identification of Chinese sentences in the adoptee group exhibited a strong relationship to the HSK score ($r = 0.83$, $p = 0.0052$), but not age of adoption ($p = 0.25$).

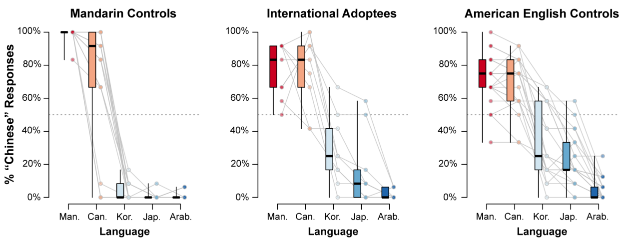

Figure 4: Percentage of sentences identified as Chinese when hearing sentences from five different languages.

3.5. FFR pitch tracking accuracy

Table 3 lists average FFR pitch tracking accuracy by contour and group (Fig. 5). We conducted independent samples t-tests to determine pairwise group differences in the pitch tracking accuracies. We found that Mandarin controls and international adoptees exhibited similar pitch tracking accuracies ($p > 0.06$). Compared to the native English group, significantly higher FFR pitch tracking accuracy for T1 was found in the Mandarin control ($t_{39.8} = 2.08$, $p = 0.047$). Interestingly, we found that compared to native English listeners, pitch tracking accuracy for T4 was significantly higher in the adoptees ($t_{18.3} = 2.30$, $p = 0.034$).

	Mandarin Controls	International Adoptees	English Controls
T1: Level	0.20 ± 0.28	0.25 ± 0.34	0.04 ± 0.17
T2: Rise	0.61 ± 0.47	0.92 ± 0.15	0.67 ± 0.49
T3: Dip	0.77 ± 0.31	0.75 ± 0.33	0.70 ± 0.35
T4: Fall	0.84 ± 0.27	0.96 ± 0.06	0.82 ± 0.25

Table 3 Mean FFR pitch tracking accuracy (Pearson's r) in the participant groups for each Mandarin tone contour.

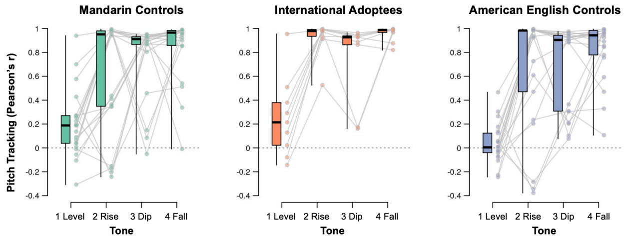

Figure 5: Participants' FFR pitch tracking accuracy by group for each Mandarin tone.

4. DISCUSSIONS

This study examined whether international adoptees maintained traces of their heritage language even after a prolonged period of limited or no exposure to it. Behaviorally, we found that the international adoptees partially maintained phonological sensitivity to pitch contours; although the international adoptees were significantly poorer at perceiving pitch compared to Mandarin controls, adoptees were significantly more accurate at pitch perception than native English controls. These results suggest that adoptees do not retain fully native-like phonological representations. However, the adoptees' superior performance to the native English group suggests that they may still have some representation of early acquired linguistic features when re-exposed to Mandarin [7,10].

We also found that higher-level linguistic knowledge (lexical/semantic) was not retained by the international adoptees as demonstrated by their poorer performances in the Chinese Stroop tasks and Chinese sentence identification compared to Mandarin controls. Given the fact that the heritage language's words are lost within months of adoption and into adulthood [7,10], it is expected that adoptees would have weaker semantic connections between the word and the associated semantic meaning.

Analysis of the FFR data yielded modest results potentially due to noise and sample size; nevertheless, the adoptee group showed generally better FFR pitch tracking than their native English-speaking peers.

While we cannot rule out influences of post-adoption Mandarin exposure on the enhanced phonological and pitch sensitivities in the adoptees vs. native English speakers, we doubt that adult language learning alone explains their better tone sensitivity, as these low-level behavioral measures were not reliably correlated to their Mandarin competence (i.e., HSK).

In sum, these results suggest that the adoptees may have latent phonological representations of their heritage language, which may support their ability to learn Mandarin [13]. These data are consistent with neural commitment theory, in which early language environments have long-lasting effects on auditory and linguistic representations.

5. REFERENCES

[1] Chinese Proficiency Test 汉语水平考试 Hanyu Shuiping Kaoshi (HSK): www.chinesetest.cn

[2] Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* 25, 161-168.

[3] Kuhl, P.K., Tsao, F.M., Liu, H.M., Zhang, Y., & De Boer, B. 2001. Language/culture/mind/brain: Progress at the margins between disciplines. *Annals of the New York Academy of Sciences* 935, 136-174.

[4] 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, 606-608.

[5] Moulines, E. & Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9, 453–467.

[6] Norman, G., Bylund, E., & Thierry, G. 2022. Irreversible specialization for speech perception in early international adoptees. *Cerebral Cortex* 32, 3777-3785.

[7] Pallier, C., Dehaene, S., Poline, J.B., LeBihan, D., Argenti, A.M., Dupoux, E., & Mehler, J. 2003. Brain imaging of language plasticity in adopted adults: Can a second language replace the first?. *Cerebral Cortex* 13, 155-161.

[8] Perrachione, T.K., Lee, J., Ha, L.Y.Y., & Wong, P.C.M. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J. Acoust. Soc. Am.* 130, 461-472.

[9] Pierce, L.J., Klein, D., Chen, J.K., Delcenserie, A., & Genesee, F. 2014. Mapping the unconscious maintenance of a lost first language. *Proc. Nat. Acad. Sci. USA.* 111, 17314-17319.

[10] Ventureyra, V.A., Pallier, C., & Yoo, H.Y. 2004. The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, 17, 79-91.

[11] Werker, J.F. & Tees, R.C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49-63.

[12] Wong, P.C.M. & Perrachione, T.K. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28, 565-585.

[13] Zhou, W., Broersma, M., & Cutler, A., 2021. Asymmetric memory for birth language perception versus production in young international adoptees. *Cognition*, 213, 104788.