Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Pitch Target Realization in Putonghua Tone Production of Children from Dialect-Speaking Regions

Mengxue Cao[1], *Tianxin Zheng*[2], *Jiewen Zheng*[2]

[1]Key Laboratory of Linguistics, Chinese Academy of Social Sciences, China
[2]School of Chinese Language and Literature, Beijing Normal University, China
caomx@cass.org.cn, zhengtianxin@mail.bnu.edu.cn, jiewen.zheng@mail.bnu.edu.cn

Abstract

Tonal production can be understood as a target realization process. In a hybrid linguistic environment, tonal targets may interfere with one another. This study examines the production of Putonghua tones from both on- and off-target perspectives, using data from 139 children (aged 35–71 months) with Changli dialect exposure, drawn from the CL-CHILD corpus. Results reveal that: (1) children's pitch target realization is universally constrained by physiological limitations, exhibiting cross-linguistic commonalities; (2) dialect exposure causes persistent pitch target interference over an extended period; (3) off-target pronunciations stem from phonetic similarities across tonal categories and emerge through mutual interference between multiple targets. These findings underscore that children's tonal development is a dynamic process of target approximation, shaped by both physiological constraints and language experience.

Index Terms: Pitch target, lexical tone, Putonghua, dialect exposure, preschool children

1. Introduction

Putonghua is a standardized language variety based on the phonological and grammatical system of Beijing Mandarin [1]. Putonghua-speaking children typically acquire lexical tones early (e. g., by 22 months) [1, 2, 3], owing to the high saliency of tones in tonal languages [1, 4, 5]. In terms of acoustic realizations, however, children aged over 36 months still have difficulties in approaching the accurate pitch targets of F_0 (i. e., fundamental frequency) range, slope, and curvature, despite their abilities to produce nearly adult-like pitch contours [6, 7, 8, 9]. The physiological limitation of children's immature motor controls is argued as the main cause for the lag between the phonological and the articulatory–acoustic development [9, 10].

Consider Chinese as a lingua franca [11], the tonal development of children from dialect-speaking regions, however, are not equally investigated. Limited observations of Putonghua tonal production of children from Gaobeidian [12] and Fuzhou [13] indicate that children with dialect exposure exhibit similar patterns, such as difficulties in approaching the low pitch target, compared to their Putonghua peers [6, 7, 8, 9]. Cross-dialect evidence showing that Fuzhou children also struggle to realize the low pitch target of tones in Fuzhou dialect [13] further highlights the shared deviant patterns among children, driven by physiological constraints [14]. Regarding the impact of dialect exposure on Putonghua production [12, 13, 15, 16], decreased pitch accuracy suggests that phonetic interference could potentially affect children's acoustic performance [17].

Within the framework of the Target Approximation model, pitch modulation can be regarded as a target approximation process, representing a goal-oriented model of F_0 movements [18]. Consequently, children's tonal productions are perceived as phonologically correct when their F_0 movements approximate the corresponding pitch targets closely enough. Compared to adults, the acoustic realization of correct pronunciations may vary in F_0 details such as pitch level or curvature, yet they are still considered on-target pronunciations [9]. When F_0 movements deviate from the original pitch targets, however, tonal productions are perceived as phonologically incorrect and classified as off-target pronunciations.

Studies on Putonghua-speaking children primarily focus on the acoustic characteristics of on-target pronunciations, while the pitch realization of off-target ones remains under-explored, with limited evidence available [9, 19]. In reality, children rarely attempt to produce incorrect tones deliberately; rather, off-target pronunciations emerge as direct outcomes of their tonal target realization strategies. Therefore, a comprehensive analysis of children's pitch target realization, encompassing both on- and off-target perspectives, is crucial.

The CL-CHILD corpus, proposed in our earlier research [20], features high-quality audio and annotation, and captures the phonological development of Putonghua in preschool children exposed to Changli dialect. It serves as a valuable resource for examining children's tonal target production in a complex linguistic environment.

The present study investigates how Changli preschool children regulate F_0 details for their pitch target realization when producing Putonghua lexical tones. Specifically, the study aims to address the following questions. First, what are the acoustic characteristics of F_0 contour in on- and off-target pronunciations? Second, what factors influence the pitch target realization process?

2. Methods

2.1. Materials

The phonological systems of Changli dialect and Putonghua are largely consistent in terms of initials and finals [21]. While they share the same tonal categories, their tonal features differ. Using a five-point scale notation [22], Changli tones are represented as T1 /41/, T2 /243/, T3 /213/, and T4 /453/ [23], whereas Putonghua tones are T1 /55/, T2 /35/, T3 /214/, and T4 /51/ [24].

The speech data of 139 Changli children (62 males, 77 females, aged 35–71 months) were adopted from the CL-CHILD corpus we previously developed [20]. The recording script included 27 monosyllabic words, comprising six tokens for T1, seven for T2, eight for T3, and six for T4. The annotation, processing, and cross-validation of the speech data were carried out in our previous study [20], where detailed procedures are described. Briefly, data were annotated in Praat [25] by two trained postgraduate students and a professional researcher.

 10.21437/Interspeech.2025-1237

Segmental boundaries and on-/off-target pronunciations were marked based on phonological perceptual judgments.

A parallel speech corpus of native Putonghua-speaking adults (20 males, 20 females, 22.5 $\pm$ 1.2 years) was recorded and annotated following the same specifications as the CL-CHILD corpus.

To examine age-related patterns, the children's data were divided into three age groups, based on developmental norms for Putonghua-speaking children in Beijing [3]: 35–47 months, 48–59 months, and 60–71 months (see Table 1).

Table 1: *Information of the 139 children in three age groups*

Age group (months)	Total	Male	Female
35–47	33	14	19
48–59	69	29	40
60–71	37	19	18

2.2. Data processing

After excluding audio samples with noticeable creaky voice quality, 3178 tokens from children and 912 tokens from adults were kept for acoustic analysis. The F_0 raw data was measured with 10 equidistant samples from the finals of each syllable using the raw auto-correlation algorithm [26], and extracted using the *Get mean (curve)* function in Praat [25]. To reduce inter-speaker differences, the F_0 raw data was first converted to semitones (with 50 Hz as reference) and then normalized using the *Lz* (i. e., logarithmic z-score) method [27] on a per-speaker basis, as shown in Equation 1:

$$Lz = (\log_{10} F_0st - \mu)/\sigma \quad (1)$$

where F_0st represents the measured F_0 value in semitones, and μ and σ denote the mean and standard deviation, respectively, of all logarithmic F_0st values for the given speaker.

To minimize the impact of extreme values on statistical efficiency, on-target and off-target pronunciations were pre-screened separately in the following procedure. First, the standard deviation of the second-order difference across the 10 data points of each pronunciation was calculated. Then, only data within 95$^{\text{th}}$ percentile of the standard deviation distribution for the corresponding tonal category were kept for further analysis.

2.3. Acoustic analysis

To investigate the pitch target realization of on-target pronunciations, acoustic differences between children of the three age groups and adults were measured. For off-target pronunciations, no distinctions between age groups were made. Instead, four groups of on-target pronunciations from children and adults were employed to compare with the off-target ones, in order to reveal the constituent nature of different pitch targets. For instance, if T2 was mispronounced as T3, the off-target pronunciation was then compared with the on-target T2 and T3 productions from both children and adults.

Qualitative analysis was conducted by examining the F_0 contours plotted with the mean and 95% confidence interval. Quantitative analysis involved the functional principal components analysis (FPCA) [28], followed by the Kruskal-Wallis test and the Dunn's test.

FPCA is a method for quantifying F_0 dynamics [28] and has been verified in studies for comparative analysis of pitch data with temporal information [29, 30]. Given an input contour $f(t)$, FPCA provides the mean curve $\mu(t)$, the principal component curves (PCs), and the weights (i. e., PC scores) which are used to scale the PCs to the mean curve $\mu(t)$, that ultimately produce the best approximation of $f(t)$. In this study, PC1 and PC2, which exhibit high explanatory power, were selected as the primary metrics for quantitative analysis.

The Kruskal-Wallis test was then applied to examine the PC scores (i. e., s1 for PC1 and s2 for PC2) across groups. Finally, post-hoc pairwise comparisons were conducted using Dunn's test, with p-value adjusted using Bonferroni's correction.

3. Results

3.1. On-target pronunciations

The analysis of on-target pronunciations focuses on the acoustic differences between each child group and the adult group, aiming to reveal the details of F_0 distinctions during target realization.

PC1 and PC2 of the FPCA modeling together account for over 80% variance of the F_0 trajectories of each tonal category (see Table 2). PC1 primarily captures variations in the overall level of the F_0 trajectory, while PC2 captures variation in F_0 trajectory shape through combined effects of peak alignment, slope, and F_0 range (see Figure 1).

Table 2: *The variance covered by PC1 and PC2 of each tone*

Tone	PC1	PC2	PC1+PC2
T1	94%	4%	98%
T2	77%	13%	90%
T3	63%	18%	81%
T4	70%	20%	90%

3.1.1. T1

For T1, while no visible difference in the F_0 contours is found between each of the three child age groups, children of all age groups exhibit overall pitch lowering and a descending tail compared to adults.

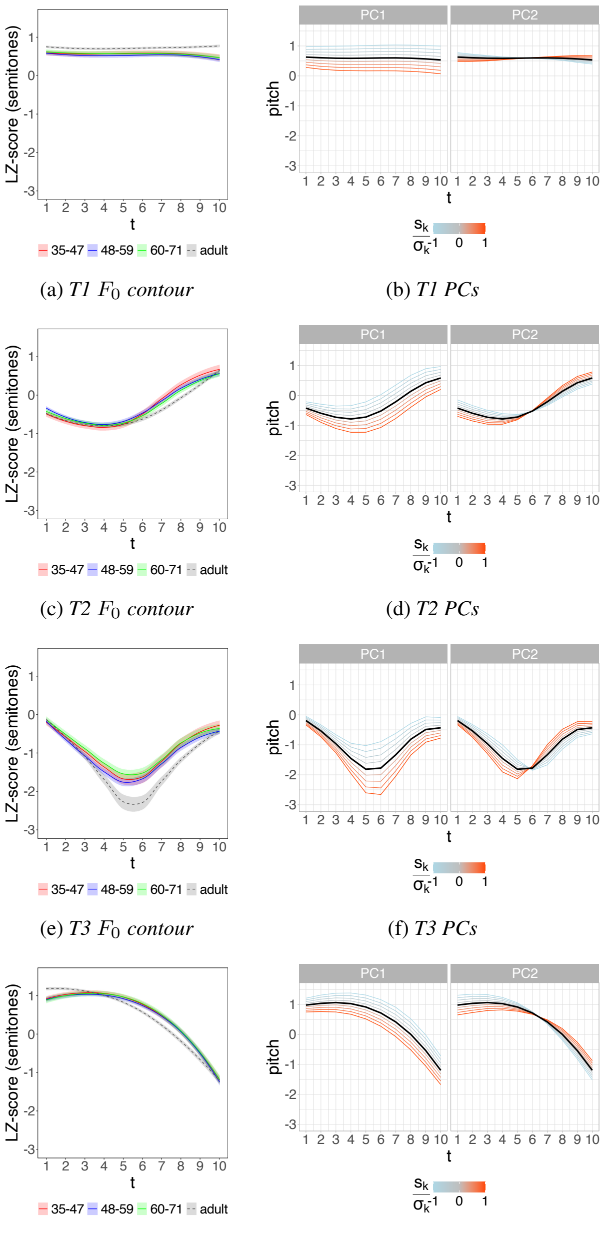
The Kruskal-Wallis test shows significant differences in s1 [$\chi^2(3) = 39.422$, $p < 0.001$] and s2 [$\chi^2(3) = 17.853$, $p < 0.001$] among child and adult groups. The Dunn's test further reveals that, while no significant differences in s1 and s2 are found between each child groups, children in all three groups (all, $p < 0.001$) show significant differences in s1, and the 48–59 ($p = 0.006$) and 60–71 ($p < 0.001$) groups show significant differences in s2, compared to adults.

The above results indicate that when producing T1, children across all age groups share a common pattern with lower pitch level and a falling pitch tail compared to the adults' realization, while no developmental changes are observed.

3.1.2. T2

For T2, Changli children of all age groups exhibit an S-shaped F_0 curvature, with a concave curve at the beginning, a rapid rise in the middle, and a notable convex curve towards the end. Compared to adults, the main differences lie in the convex portion of the F_0 trajectory.

The Kruskal-Wallis test shows significant differences in s1 [$\chi^2(3) = 10.202$, $p = 0.017$] and s2 [$\chi^2(3) = 23.738$, $p <$



(a) *T1 F_0 contour* (b) *T1 PCs*

(c) *T2 F_0 contour* (d) *T2 PCs*

(e) *T3 F_0 contour* (f) *T3 PCs*

(g) *T4 F_0 contour* (h) *T4 PCs*

Figure 1: *F_0 contour and PCs of on-target pronunciations. The F_0 contours reflect actual pitch movements, while the PCs represent modeled variation patterns. In the PC plots, the bold black curve show the mean curves $\mu(t)$, and the red and blue curves represent $\mu(t)$ plus or minus multiple standard deviations along selected PCs, illustrating the effect of PC scores on pitch contour variation.*

0.001] among child and adult groups. The Dunn's test further reveals that significant differences can only be found between the 48–59 group and adults ($p = 0.016$) in s1, and between the 35–47 group and the 48–59 ($p < 0.001$), the 60–71 ($p = 0.016$), and adults ($p < 0.001$) in s2.

The results indicate that the production of T2 in Changli children exhibits varying degrees of curvature precision development between 35 and 71 months. As age increases, the S-shaped tendency gradually reduces, and the tail curvature approaches closer to the adult target, with a significant improvement towards adult-like level around 48–59 months. While manipulating the curvature target, however, the F_0 level rises during the same period. Only in the 60–71 group, do children gradually balance F_0 level and contour targets, closely approximating adults' production. The above developmental pattern reflects a series of spiraling stages in the refinement of tonal target realization.

3.1.3. T3

For T3, Changli children of all age groups share a similar F_0 pattern, featuring a concave contour close to adults. However, a notable difference remains between children and adults, as children constantly exhibit insufficient concavity.

The Kruskal-Wallis test shows significant differences only in s1 [$\chi^2(3) = 50.401$, $p < 0.001$] among child and adult groups. The Dunn's test further reveals that, while no significant differences in s1 and s2 are found between each child group, children in all three groups (all, $p < 0.001$) differ significantly from adults in s1 only.

The results suggest that pitch accuracy in T3 production displays no significant improvement among Changli children between 35 and 71 months. Although they acquire the contour features of T3 relatively well, their control over low pitch targets remains underdeveloped, leading to insufficient concavity of the F_0 curve compared to adults.

3.1.4. T4

For T4, while no noticeable difference in the F_0 contours are observed between each of the three child age groups, children generally exhibit a convex contour with a slow rise at the start compared to adults.

The Kruskal-Wallis test shows significant differences in s1 [$\chi^2(3) = 15.204$, $p = 0.002$] and s2 [$\chi^2(3) = 42.886$, $p < 0.001$] among child and adult groups. The Dunn's test further reveals that, while no significant differences in s1 and s2 are found between each child group, children in all three groups differ significantly from adults in s1 (35–47, $p = 0.022$; 48–59, $p = 0.007$; 60–71, $p = 0.006$) and s2 (all, $p < 0.001$).

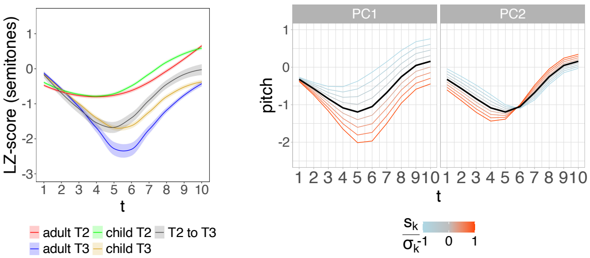
These results indicate that the convex contour pattern in T4 production remains stable across all age groups, suggesting a persistent deviation from the adult target without gradual refinement over time.

3.2. Off-target pronunciations

The error pattern with an error rate over 10% (i. e., T2 mispronounced as T3 at 16%) identified in our previous study [20] was selected for investigation. The analysis of off-target pronunciations focuses on the acoustic commonalities between the off-target group and the four on-target groups, aiming to uncover the underlying targets contributing to the deviated surface F_0 contour during pitch target realization.

PC1 and PC2, which capture the same variation components as those in the on-target analysis, together account for 89% (PC1, 77%; PC2, 12%) variance of the F_0 trajectories.

Clear differences of the F_0 contours can be observed between each of the 5 groups (see Figure 2). The off-target tone (i. e., T2 to T3) deviates from the T2 of both child and adult groups, exhibiting an F_0 level and descending portion similar to the on-target T3 in children, while its convex curve toward the end mimics the pitch pattern of the on-target T2 in children.



(a) F_0 *contour* (b) *PCs*

Figure 2: F_0 *contour and PCs of the off-target pronunciations.*

The Kruskal-Wallis test shows significant differences in s1 [$\chi^2(4) = 820.77$, $p < 0.001$] and s2 [$\chi^2(4) = 135.76$, $p < 0.001$] among child and adult groups. The Dunn's test further reveals that the off-target tone is similar in F_0 level (i. e., s1) to the on-target T3 in children ($p = 0.418$) and in F_0 shape (i. e., s2) to the on-target T2 in both children ($p = 1.000$) and adults ($p = 1.000$).

These complex acoustic characteristics indicate that the off-target pronunciation differs significantly from the on-target T2 and T3 in both children and adults. It suggests that the off-target pronunciation, perceived as T3 rather than its original category T2, represents a blend of pitch targets from both tones, leading to a mixed pitch realization in shaping the surface F_0 contour.

4. Discussion

4.1. Constraints of physiological development

During the production of Putonghua tones, little developmental evidence is observed across different age groups in Changli children, except for T2. Although on-target pronunciations are perceived as phonologically correct, noticeable differences in F_0 contours between children and adults persist across all four tonal categories. It indicates a common articulatory–acoustic gap between children and adults during tonal realization.

Patterns observed in Changli children are also reported in their Putonghua-speaking peers [6, 7, 8]. For instance, Putonghua children display similar S-shaped contours in T2 and show difficulties in approaching the low target in T3 [7, 9, 12]. These cross-dialect evidences suggest that preschool children suffer common difficulties in precise pitch control in both F_0 level and shape, regardless of their language environment.

The above difficulties are evidently linked to the physiological development constraints of vocal organs [9, 10]. Moreover, off-target pronunciations share more acoustic features with children than adults, suggesting a coherent pattern caused by children's physiological constraints.

4.2. Effects of dialect exposure

Children with dialect exposure often exhibit interference in pitch target realization, leading to decreased pitch accuracy compared to their Putonghua-speaking peers [12, 13, 15].

The F_0 surface of Putonghua tones produced by Changli children exhibits a falling tail in T1 and a convex contour in T4, mirroring the corresponding tonal features of Changli dialect in certain tonal categories [23]. In addition, the convex tail in T2 and the higher F_0 level in T3 are also linked to the competition between T2 and T3 in Changli dialect [31]. It is particularly evident as Putonghua-speaking children achieve adult-like T3 production by 48 months [8], whereas Changli children between 35 and 71 months consistently deviate from the adult target. The same phenomenon is observed in children aged 83–90 months with Fuzhou dialect exposure [13], indicating a prolonged impact of dialect influence on tonal development. Moreover, dialect features are also present in off-target pronunciations, suggesting a universal pitch realization strategy among children with dialect exposure.

Children exposed to multiple languages often exhibit a compromise in acoustic features compared to their monolingual peers, achieving acceptable phonetic realizations but with less precise acoustic targets [32, 33, 34, 35]. Consequently, under the influence of dialectal tonal features, children continue to exhibit dialectal characteristics in their production of both on- and off-target Putonghua tones for an extended period.

4.3. Interference of multiple tonal targets

In both Changli dialect [23, 31] and Putonghua [10], T2 and T3 are more likely to be confused in both production and perception due to their similar pitch targets. As predicted by the Perceptual Assimilation Model (PAM) [36], children tend to merge categories with similar phonetic features, resulting in deviations and a blending of phonetic targets. Consequently, when attempting to produce Putonghua T2, Changli children are more likely to deviate in their acoustic realization from the original T2 target, blending it with features of the T3 target in their off-target pronunciations, resulting an incomplete assimilation.

Those deviations are more influenced by phonetic similarities than by direct confusion between phonological categories. As a result, the off-target pronunciations do not fully conform to the pitch targets of the deviated tonal categories, but instead exhibits a blend of acoustic features from both the original and deviated ones. In the case of the relation between T2 and T3, particularly in Changli children, the F_0 surface of the off-target tone is shaped by the F_0 level of the deviated tonal category, along with the curvature details (e. g., peak alignment, slope, and F_0 range) of the original one.

5. Conclusion

This study demonstrates that Putonghua tonal production in children with dialect exposure can be understood as a target realization process, shaped by both physiological constraints and language experiences. Off-target pronunciations should be regarded as a specific case of pitch target realization, where multiple pitch targets from different tonal categories interact due to assimilation driven by phonetic similarities.

6. Acknowledgements

This study was supported by the Key Laboratory of Linguistics, Chinese Academy of Social Sciences (Project No. 2024SYZH001).

7. References

[1] H. Zhu, *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Multilingual Matters, 2002.

[2] W. Li, H. Zhu, D. Barbara, T. Jiang, D. Peng, and H. Shu, "Phonological acquisition of Putonghua-speaking children," *Acta Psychologica Sinica*, vol. 32, no. 02, pp. 170–176, 2000.

[3] J. Gao, *The Manual of the Articulation Test for 1.5- to 6-year-old Putonghua-speaking Children in Beijing*. Beijing: China Social Sciences Press, 2023.

[4] J. F. Werker and S. Curtin, "PRIMIR: A developmental framework of infant speech processing," *Language Learning and Development*, vol. 1, no. 2, pp. 197–234, 2005.

[5] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1493, pp. 979–1000, 2008.

[6] N. X. Rattanasone, P. Tang, I. Yuen, L. Gao, and K. Demuth, "3-year-olds produce pitch contours consistent with Mandarin tone 3 sandhi," in *Proc. Sixteenth Australasian International Conference on Speech Science and Technology*, 2016, pp. 5–8.

[7] ——, "Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like," *Frontiers in Psychology*, vol. 9, p. 817, 2018.

[8] P. Tang, I. Yuen, N. X. Rattanasone, L. Gao, and K. Demuth, "The acquisition of phonological alternations: The case of the Mandarin tone sandhi process," *Applied Psycholinguistics*, vol. 40, no. 6, pp. 1495–1526, 2019.

[9] P. Wong, "Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions," *Journal of Phonetics*, vol. 40, no. 1, pp. 141–151, 2012.

[10] P. Wong, R. G. Schwartz, and J. J. Jenkins, "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 5, pp. 1065–1079, 2005.

[11] D. C. Li, "Chinese as a lingua franca in greater China," *Annual Review of Applied Linguistics*, vol. 26, pp. 149–176, 2006.

[12] Y. Liu and Z. Meng, "A survey on Mandarin tone patterns of primary and secondary school students in Gaobeidian area," in *Proc. PCC 2012 – 10th Phonetic Conference of China*, 2012, pp. 191–195.

[13] S. You, Y. Du, and Q. Chen, "Production of Mandarin and Fuzhou lexical tones in six-to seven-year-old Mandarin–Fuzhou bilingual children," *Acta Linguistica Academica*, vol. 71, no. 3, pp. 309–337, 2024.

[14] A. Holm and B. Dodd, "Phonological development and disorder of bilingual children acquiring Cantonese and English," *Phonological Development and Disorders in Children: A Multilingual Perspective*, pp. 286–325, 2006.

[15] Y. Liu, J. Guo, and Z. Meng, "Perception and production of Mandarin tones by pupils in Xinjiang ethnic areas," *Journal of Tsinghua University (Science and Technology)*, vol. 6, pp. 823–827, 2013.

[16] L. So and C. Leung, "Phonological development of Cantonese-Putonghua bilingual children," *Phonological Development and Disorders in Children: A Multilingual Perspective*, pp. 413–428, 2006.

[17] R. Ellis, *Second Language Acquisition*. New York: Oxford University Press, 1997.

[18] Y. Xu and Q. E. Wang, "Pitch targets and their realization: Evidence from Mandarin Chinese," *Speech Communication*, vol. 33, no. 4, pp. 319–337, 2001.

[19] Y. Lin, "Learning to speak a minority language at school: Mandarin speech development of children in a Chinese–English bilingual program in Canada," Ph.D. dissertation, University of Alberta, 2023.

[20] J. Zheng, T. Zheng, and M. Cao, "CL-CHILD corpus: The phonological development of Putonghua in children from dialect-speaking regions," in *Proc. O-COCOSDA 2024 – 27th Conference of the Oriental COCOSDA International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques*. IEEE, 2024, pp. 1–6.

[21] M. Cao, "Insight into the phonetic features of Changli-Town dialect," *Chinese Journal of Phonetics*, vol. 13, no. 1, pp. 81–89, 2020.

[22] Y.-R. Chao, "A system of tone letters," *Le Maître Phonétique*, vol. 45, no. 30, pp. 24–27, 1930.

[23] M. Cao, "Study of Changli-Town dialect's citation tones," *Chinese Journal of Phonetics*, vol. 8, no. 1, pp. 54–60, 2017.

[24] Y.-R. Chao, *A Grammar of Spoken Chinese*. University of California Press, 1968.

[25] B. Paul and W. David, *Praat: Doing Phonetics by Computer*, 2024. [Online]. Available: http://www.praat.org/

[26] P. Boersma *et al.*, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proc. the Institute of Phonetic Sciences*, vol. 17, no. 1193. Amsterdam, 1993, pp. 97–110.

[27] X. Zhu, "F0 normalization: How to deal with between-speaker tonal variations?" *Linguistic Sciences*, vol. 9, pp. 3–19, 2004.

[28] M. Gubian, F. Torreira, and L. Boves, "Using functional data analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, pp. 16–40, 2015.

[29] J. Cole, J. Steffman, and A. Awwad, "Functional modeling of f0 variation across speakers and between phonological categories: Rising pitch accents in American English," in *Proc. SpeechProsody 2024*, 2024, pp. 1020–1024.

[30] D. El Zarka, A. Kelterer, M. Gubian, and B. Schuppler, "The prosody of theme, rheme and focus in Egyptian Arabic: A quantitative investigation of tunes, configurations and speaker variability," *Speech Communication*, vol. 160, p. 103082, 2024.

[31] M. Cao, "The tone sandhi and neutral tone of disyllabic non-duplicative words in Changli dialect in the town of Changli county of Hebei province," *Dialect*, vol. 42, no. 04, pp. 401–410, 2020.

[32] A. Caramazza, G. H. Yeni-Komshian, E. B. Zurif, and E. Carbone, "The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals," *The Journal of the Acoustical Society of America*, vol. 54, no. 2, pp. 421–428, 1973.

[33] I. Watson, "Phonological processing in two languages," *Language Processing in Bilingual Children*, pp. 25–48, 1991.

[34] J. Yang and R. A. Fox, "L1–L2 interactions of vowel systems in young bilingual Mandarin–English children," *Journal of Phonetics*, vol. 65, pp. 60–76, 2017.

[35] J. Yang, "Comparison of VOTs in Mandarin–English bilingual children and corresponding monolingual children and adults," *Second Language Research*, vol. 37, no. 1, pp. 3–26, 2021.

[36] C. T. Best *et al.*, "The emergence of native-language phonological influences in infants: A perceptual assimilation model," *The Development of Speech Perception: The Transition From Speech Sounds to Spoken Words*, vol. 167, no. 224, pp. 233–277, 1994.