*Journal of Child Language* (2020), 1–12
doi:10.1017/S0305000920000239

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

# Going beyond F0: The acquisition of Mandarin tones

Nari RHEE[1]*, Aoju CHEN[2], and Jianjing KUANG[1]*

[1]University of Pennsylvania, USA and [2]Utrecht University, Netherlands
*Corresponding authors: Department of Linguistics, 3401-C Walnut Street, Suite 300, C-Wing, University of Pennsylvania, Philadelphia, PA 19104-6228, USA E-mails: nrhee@sas.upenn.edu and kuangj@ling.upenn.edu

(Received 9 July 2019; revised 13 January 2020; accepted 7 April 2020)

**Abstract**

Using a semi-spontaneous speech corpus, we present evidence from computational modelling of tonal productions from Mandarin-speaking children (4- to 11-years old) and adults, showing that children exceed the adult-level tonal distinction at the age of 7 to 8 years using F0 cues, but do not reach the high adult-level distinction using spectral cues even at the age of 10 to 11 years. The difference in the developmental curves of F0 and spectral cues suggests that, in Mandarin tone production, secondary cues continue to develop even after the mastery of primary cues.

**Keywords:** tone acquisition; spectral cues; Mandarin

## Introduction

The development of a child's linguistic abilities spans the entire childhood and even into early adolescence. While the acquisition of native phonological contrasts begins in infancy, children who are capable of perceiving or producing the categories may not have reached fully adult-like competence in all associated cues until much later in the development. This study investigates the acquisition of tones in a tonal language, Mandarin. The four tonal categories of Mandarin (T1: high level, T2: mid-rising, T3: low-dipping, and T4: falling) are primarily distinguished by height and shape of the fundamental frequency (F0). However, further investigation of the tones has revealed that, in addition to F0, voice quality also serves as an important cue in the production and perception of Mandarin tonal contrast for adults (Belotel-Grenié & Grenié, 1994). Previous studies have examined the development of Mandarin tonal contrasts both in production (e.g., Hua & Dodd, 2000; Li & Thompson, 1977; Wong, 2013) and perception (e.g., F. Chen, Peng, Yan & Wang, 2017; Yeung, Chen & Werker, 2013), yet studies have been predominantly focused on the development of F0 distinctions in children under the age of 7 years. Hence, little is known about the full developmental trajectory of the tonal contrasts and all their associated cues. This study investigates both F0 and voice quality cues in the tonal production of children from a wider range of ages (4 to 11), to contribute to a better understanding of the full developmental trajectory for Mandarin tone

production, and how physiologically correlated cues become linguistically useful during the process of language development.

### *Mastery of linguistic features*

Children's development of speech production and perception is restricted by physiological factors such as premature vocal tract and motor control (Tingley & Allen, 1975), as well as cognitive factors such as auditory-feedback mapping (P. Liu, Chen, Larson, Huang & Liu, 2010; Shiller, Gracco & Rvachew, 2010), which continue to develop throughout childhood and early adolescence. Studies on English vowel production have reported that younger children's productions of monophthongal vowels have higher variability than those of older children or adults; adult-level stability is achieved at the age of 12 years for durational cues and at around 14 and 15 for F0, formants, and spectral envelope (Lee, Potamianos & Narayanan, 1999), and the vowel space area of 14-year-olds is larger than that of adults (Pettinato, Tuomainen, Granlund & Hazan, 2016). Such continuous development across childhood has also been observed in the acquisition of phrase-level prosody (see Chen, Esteve-Gibert, Prieto & Redford, unpublished observations). For example, English-learning children can produce distinctive global rising and falling F0 contours at 17 months (Galligan, 1987), but full mastery of the rising contours is not achieved even at the age of 4 years (Patel & Brayton, 2009). Similar latency in the mastery of phrase-level prosody has been found in Spanish-speaking children, who are not able to produce adult-like F0 scaling (i.e., the difference between the highest and lowest F0) in phrase-final falling F0 contour until the age of 6 years (Astruc, Payne, Post, Vanrell & Prieto, 2013) and German-speaking children, who are not yet adult-like in F0 scaling and alignment (i.e., the timing of the highest and lowest F0) in phrase-final rising F0 contour even by the age of 7 years (De Ruiter, 2010).

More specifically, different prosodic cues are mastered at different stages of language development. For example, durational cues of English prosodic contrasts are reliably produced at the age of 4 years, while other cues (such as amplitude and F0) are produced at a later age (Patel & Brayton, 2009).

This paper expands the study of the developmental trajectory of linguistic features to first language acquisition of tones, investigating the mastery of the tonal contrasts using different cues.

### *The acquisition of Mandarin tonal contrast*

Past studies have examined the development of tones in the production and perception of Mandarin-learning infants and children. In perception, early sensitivity to lexical tones is observed at as early as 4 months of age (Yeung *et al.*, 2013), but the boundaries of Mandarin-learning children's categorical perception of T1 and T2 are not as sharp as adults until the age of 6 years, and the discrimination accuracy does not reach the adult level even at the age of 7 years (F. Chen *et al.*, 2017).

Studies have presented contradicting results regarding the developmental trajectory of tone production in Mandarin. Some studies have suggested that children become capable of contrasting tones at a high level of accuracy before the age of 3 years (Hua & Dodd, 2000; Li & Thompson, 1977). However, these studies have relied on one Mandarin speaker to determine the accuracy of the production. Wong (2013) had multiple native Mandarin speakers identify tones in children's tonal production after removing the

segmental information by low-pass filtering. She has found the tone identification accuracy in the children's production is not comparable to that in the adults' production even by the age of 5 years for monosyllabic words (Wong, 2013) and by the age of 6 years for disyllabic words (Wong & Strange, 2017). In sum, studies together suggest that though children produce some tonal contrasts as early as at age 3, their production, even at 5 to 6 years of age, is more variable and less robust to the loss of high frequency segmental information than adults' tonal production.

### Voice quality in tones

Mandarin primarily uses the F0 height and contour to distinguish the four tones (Yip, 2002). Nevertheless, tonal contrasts in Mandarin are produced with several secondary cues, including duration, amplitude, and voice quality. It is well-known that, among the four lexical tones in Mandarin, the low-dipping tone (T3) is often produced with creaky voice, and this allophonic non-modal voice in turn can facilitate the perception of T3 (Belotel-Grenié & Grenié, 1994). Furthermore, the integration of voice quality cues is not unique to T3. Rather, all Mandarin tones are subject to the same phonetic mechanism (Kuang, 2017), whereby the presence of allophonic nonmodal voice is largely driven by extreme F0 targets in the sense that high F0 (e.g., in T1) is naturally associated with tense voice and low F0 (e.g., fall of T4) with creaky voice (Sundberg, 1994). As a consequence of the systematic and continuous co-variation between voice quality and F0 in tone production, the acoustic correlates of voice quality, such as spectral cues, also systematically and continuously co-vary with F0 for a given speaker (Kuang, 2017). Past studies have established that voice quality can be acoustically characterized in the spectrum, using measures such as Cepstral Peak Prominence (CPP), a measure of aperiodicity in the signal, and relative amplitude differences of the lower- and higher-frequency harmonics (e.g., H1-H2: amplitude difference between the first two harmonics; H1-A1, H1-A2, H1-A3: amplitude differences between the fundamental and the first three formants) (e.g., Keating, Esposito, Garellek, Khan & Kuang, 2011; Kreiman, Gerratt & Antoñanzas-Barroso, 2007). In Mandarin, native speakers exhibit heightened sensitivity to (Kreiman & Gerratt, 2010) and systematic use of (Keating & Esposito, 2007; Kuang, 2017) spectral cues such as H1-H2. More generally, spectral cues play an important role in pitch perception: manipulating spectral cues can significantly shift the perception of pitch height (Kuang & Liberman, 2018).

Because F0 and spectral cues are closely related both in production and perception, spectral cues themselves can thus be fairly informative in contrasting and recognizing tones. Indeed, it has been reported that tonal categories are more successfully recognized using spectral information (mel-frequency cepstral coefficients) than using F0 by a deep-neutral network classifier (Ryant, Yuan & Liberman, 2014). Human listeners of Mandarin are also able to identify tonal categories fairly accurately in the absence of F0 information, using temporal and spectral envelope cues (Kong & Zeng, 2006; S. Liu & Samuel, 2004; Whalen & Xu, 1992), highlighting the salient role of cues beyond F0.

### The current study

Despite decades of research on Mandarin tone acquisition, a few issues remain unaddressed. First, there has been no research on when tonal development reaches

the level of adults. All existent production studies, to our knowledge, have been restricted to children under the age of 6 years, and have conclusively suggested that children are still incapable of producing fully-adultlike tonal contrasts by that age. Secondly, previous studies have focused on the acquisition of F0, despite evidence for other informative co-varying cues such as voice quality. The full mastery of adult-like tonal contrast involves not only the mastery of the primary cues (F0), but also the mastery of other cues such as voice quality. It is therefore important to explore the development of the integration of the voice quality cues (such as spectral cues) in tonal production, for a better understanding of the full tone development trajectory. More specifically, studying the integration of spectral cues in tones will allow us to answer the question as to when such physiologically correlated cues become more linguistically useful.

The present study serves as the first step towards addressing these issues, by modeling the tonal spaces of children of a wider age range and of adults. We use a semi-spontaneous speech corpus which consists of SVO Mandarin sentences elicited in an interactive setting from children aged 4 to 11 years and adults (Yang & Chen, 2018). In the corpus, both the tones of the target words (i.e., the verbs) and the tones of the preceding and following words are controlled for. Also, the prosodic structure is varied at the sentence-level via manipulations of focus conditions. Because Mandarin marks focus phonetically using cues such as duration, F0 height, and F0 range (Xu, 1999; Yang & Chen, 2018), varying the focus conditions increases the variability of the tonal production, which in turn requires more intricate control over the tonal productions. The design of the corpus thus enables us to closely examine how children and adults encode the tonal contrasts, even in contexts of possible interaction and interference from a higher level of prosodic structure. This study explores the contrastivity of the tonal categories, using the focus conditions as a way to explore the maximized variability of the tones.

## Method

### *The corpus*

The corpus consists of 2969 Mandarin SVO sentences elicited as responses to either a wh-question or a comment in five focus conditions (i.e., narrow focus on the subject, verb, or object, broad focus on the whole sentence and contrastive focus on the verb) via a picture-matching game from 46 native speakers of Mandarin in four age groups: Age 4–5 (N = 12; average 5;2, range 4;6–5;10), Age 7–8 (N = 10; average 7;10, range 7;2–8;3), Age 10–11 (N = 12; average 10;9, range 10;1–11;5), and a control-group of adult speakers (N = 12; average 19 years, range 18–20 years) (Yang & Chen, 2018). One-hundred-and-sixty SVO sentences are embedded in the game. The 160 SVO sentences are unique combinations of four disyllabic subject NPs starting with the word xiao3 'little' (one noun per tone regarding the second word), eight monosyllabic verbs (one verb per lexical tone per group), and eight monosyllabic object-nouns (one noun per lexical tone per group) (Table 1). Each group-1 verb is combined once with each group-1 object-noun and each group-2 verb with each group-2 object-noun, resulting in 32 unique VPs. Each VP occurs in each of the five focus conditions, resulting in 160 VPs. The subject-nouns are approximately evenly distributed over the VPs, forming 160 SVO sentences. This procedure has made sure that in each focus condition, each

**Table 1** Words that occurred in the sentences

| | T1 | T2 | T3 | T4 |
|---|---|---|---|---|
| Subjects | 小猫<br>xiao3 mao1<br>'little cat' | 小熊<br>xiao3 xiong2<br>'little bear' | 小狗<br>xiao3 gou3<br>'little dog' | 小兔<br>xiao3 tu4<br>'little rabbit' |
| Group-1 verbs | 扔<br>reng1<br>'throw' | 埋<br>mai2<br>'bury' | 剪<br>jian3<br>'cut' | 运<br>yun4<br>'transport' |
| Group-2 verbs | 浇<br>jiao1<br>'water' | 闻<br>wen2<br>'smell' | 舔<br>tian3<br>'lick' | 卖<br>mai4<br>'sell' |
| Group-1 objects | 书<br>shu1<br>'book' | 球<br>qiu2<br>'ball' | 笔<br>bi3<br>'pen' | 菜<br>cai4<br>'vegetable' |
| Group-2 objects | 花<br>hua1<br>'flower' | 梨<br>li2<br>'pear' | 草<br>cao3<br>'grass' | 树<br>shu4<br>'tree' |

*Note*: Adopted from Yang and Chen (2018), Table 1

tone in the verbs is combined with each tone in the preceding subject-noun and with each tone in the following object-noun. The 160 sentences are divided into two lists of 80 sentences representing all verb-object tonal combinations and focus conditions, each elicited from approximately half of the participants in each age group. Only usable sentences (i.e., sentences that were produced without disfluency and in the intended context) are included in the corpus (59% for Age 4–5, 82% for Age 7–8, 90% for Age 10–11, 92% for adults) (see Yang & Chen, 2018 for exclusion criteria).

The corpus was forced-aligned using the Mandarin forced-aligner (Yuan, Ryant & Liberman, 2014). Only the sentence-medial verb syllable was used in the analysis. F0 and spectral cues were extracted using VoiceSauce (Shue, Keating, Vicenik & Yu, 2011) at 9 equidistant subsegments to yield time-normalized measurement series. Measurements from the first 3 subsegments were removed to eliminate the influence of the onset consonants. The measures used in the analysis are: STRAIGHT F0 (Kawahara, Masuda-Katsuse & De Cheveigne, 1999), CPP, and relative amplitude differences of the lower and higher harmonics (H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, H4*-2K*, 2K*-5K*, corrected for the influence of formant frequencies and bandwidths on the harmonics (Iseli, Shue & Alwan, 2007)). Tokens with F0 tracking errors (F0 jumps of greater than 50Hz between two consecutive subsegments) were removed, leaving a total of 2622 tokens for the present analysis. All extracted measures were normalized by speaker and recording session, using min-max normalization to scale between 0 and 1.

### *Computational modeling*

The tonal production of all age groups was modelled using the following sets of cues: (i) only F0, (ii) only spectral cues, and (iii) both F0 and spectral cues.

First, non-metric multidimensional scaling (MDS) was used to calculate the dissimilarity of the tonal categories in the five focus conditions, using the metaMDS function (Oksanen, Blanchet, Friendly, Kindt, Legendre, McGlinn, Minchin, O'Hara, Simpson, Solymos, Stevens, Szoecs & Wagner, 2018) in R, version 3.5.1 (R Core Team, 2018). MDS transforms the multidimensional and highly correlated acoustic space into a more interpretable, low-dimensional space. The results of the MDS were plotted to provide visual inspection of the acoustic variability within each phonological tonal category and of the overlap between categories. Moreover, to cross-verify the patterns from MDS and to quantify the extent to which each cue is objectively informative as a predictor of the tonal categories, automatic tone classification was performed using multiple machine-learning classification algorithms, namely Linear Discriminant Analysis (LDA, MASS package; Venables & Ripley, 2002), Support Vector Machine (SVM, using radial basis function kernel, e1071 package; Meyer, Dimitriadou, Hornik, Weingessel & Leisch, 2018), and Random Forest (randomforest package with default parameters; Liaw & Wiener, 2002). Average classification accuracy of tonal classification was calculated for each age group and cue set, from 100 trials of 10-fold cross-validation (see Online Supplement for more details).

## Results

### *Multidimensional scaling*

Results of MDS with two dimensions are shown in Figure 1. All plots have *stress* < 0.12, indicating that 2 dimensions were sufficient to achieve a fair fit (stress ⩽ 0.2) of the data. Overall, across all age groups, the MDS tonal spaces represented the distinction of high vs. low tone on the first dimension (x-axis), and the distinction of rising vs. falling contour on the second (y-axis).

Upon visual inspection, while even the youngest age group (Age 4–5) achieved a decent separation of the tonal categories using F0 cues (top panel of Figure 1), some improvement in the organization of the tonal clusters was observed between ages 4–5 and 7–8. For the older children (Age 7–8 and Age 10–11), though their tonal categories were not as tightly clustered as those of adults, the tones were well-organized and maximally spaced out in the scaled F0 space. In fact, the dispersion of the tonal categories in the space is larger for the older children than for adults.

In comparison to the MDS spaces based on F0, the MDS spaces based only on spectral cues (middle panel of Figure 1) exhibited a more dramatic developmental pattern in which the overall separability of the tonal categories clearly improved with speaker age. Using only spectral cues, at ages 4–5, only T3 was evidently distinguished from the other tones. In particular, for this group, the contour tones (T2 and T4) were highly overlapping in the spectral space (cf. F0 space). Children of ages 7–8 exhibited a substantial improvement in the separation of all four tones using spectral cues, with smaller overlaps between the tonal categories. The tonal categories continued to form tighter clusters in the production of the oldest children (Age 10–11), yet the adult-level separability of the tonal clusters using spectral cues was not achieved at this age range. For the adults, spectral cues alone were sufficient to distinguish all four tones.

For adults, the MDS space using both F0 and spectral cues (bottom panel of Figure 1) was similar to the MDS space using only spectral cues; using both F0 and

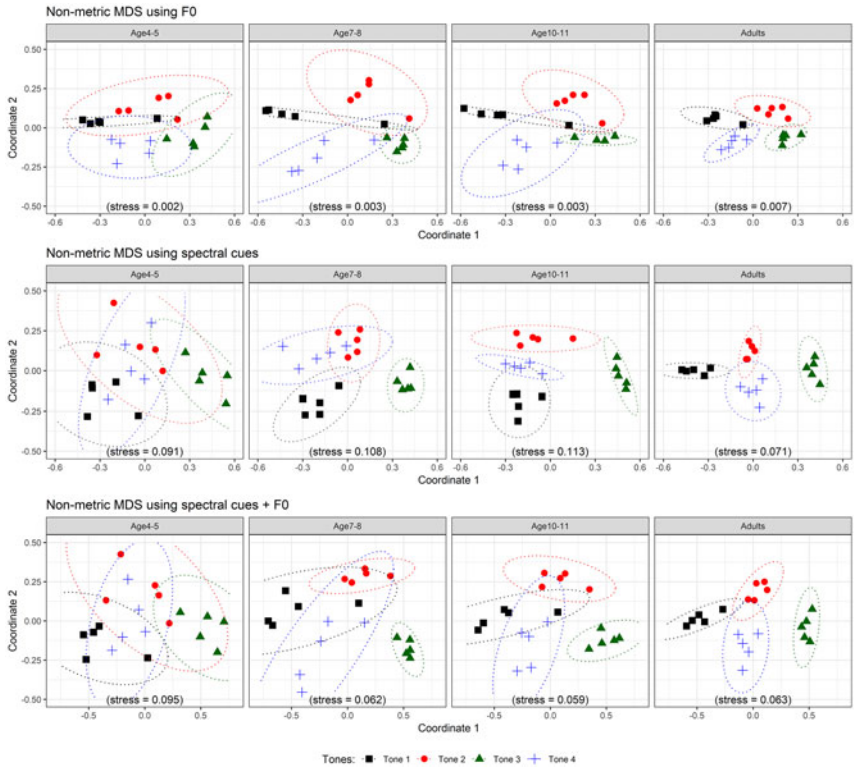


**Figure 1.** Non-metric MDS plots using only F0, only spectral cues, and both F0 and spectral cues. Ellipses are drawn for each tonal category at 95% confidence level. Each point represents tones in the five different focus conditions. The four tones are represented in different colors (T1: black square, T2: red circle, T3: green triangle, T4: blue cross).

spectral cues did not lead to an improvement in the tonal separability from using only spectral cues. In contrast, for the children, using both spectral and F0 cues achieved clearly better separation of the tonal categories than from using just spectral cues, showing that the role of F0 and spectral cues in tonal distinction are different in the production of children and adults.

Overall, results illustrated the differences in the development of F0 cues and spectral cues. F0 cues exhibited a pattern of improvement between the age of 4–5 and 7–8, achieving maximal dispersion of the tonal categories in the phonetic space at the age of 7–8, beyond the adult-level. Adults, while producing fewer between-category F0 distinctions among tones than the older children, exhibited smallest within-category variances. The integration of spectral cues in tone production exhibited a sharp development between the age of 4–5 and 7–8, and a more gradual development from the age of 7–8, particularly for tones T1, T2, and T4. Notably, adult-level contrastivity using spectral cues was not achieved by even the oldest children, who were capable of contrasting them with respect to their F0 cues. Hence, results revealed a split in the developmental curve of F0 cues and of spectral cues at the age of 7–8, when F0 cues are mastered, but spectral cues are yet to be fully learned.

### *Machine learning classification*

Accuracy rates of tonal classification using three algorithms (LDA, SVM, and Random Forest) are summarized in Figure 2. All three algorithms yielded consistent results. Using just F0 (red solid line), the highest classification accuracy was achieved for oldest children's production data (i.e., 71% for Age 10–11, Random Forest). The classification accuracy using just F0 cues was in fact lower for the Adults group than for Age 10–11 group (Random Forest), or even Age 7–8 (LDA and SVM), leading to the conclusion that older children exhibited maximally distinct tonal contrasts with F0 cues, beyond the level of adults.

Using only spectral cues (green dashed line in Figure 2), the highest accuracy was achieved for the adult data (78%, SVM). Results displayed a continuous increase in classification accuracy with age, from 54% at Age 4–5, 71% at Age 7–8, 74% at Age 10–11, and 78% for the Adults (SVM). Furthermore, the split in the developmental curves of F0 cues and spectral cues was also evident: between the age of 4–5 and 7–8, a large jump in accuracy was observed for either cue sets, but, from the age of 7–8, while tones were classified at or beyond the adult-level accuracy using only F0 cues, classification accuracy using only spectral cues continued to increase into adulthood.

Among the three cue sets (only F0; only spectral; and both), using both F0 and spectral cues (blue dotted line in Figure 2) achieved the best accuracy within every age group, which suggested that, at all ages, spectral cues were to some extent additionally useful in producing tonal contrasts. In particular, the Adults group achieved the best accuracy of 83% (SVM) when both F0 and spectral cues were used, despite having lower accuracy than the older children when just F0 cues were used.

## Discussion

This study investigated the use of F0 and spectral cues in the tonal production of Mandarin-speaking children as well as adults. The results of both MDS and machine learning classification corroborated the finding that even after children acquire the basic F0 contrasts, the contrastiveness of the tonal categories continues to improve, through integration of spectral cues in tonal production.

F0 cues were quite developed even at the age of 4–5. Nonetheless, we observed an improvement in F0 tonal contrastivity until the age of 7–8, when their F0 cues reached or even exceeded adult-level tonal contrastivity, through maximal dispersion of the F0 tonal space (MDS). This developmental pattern in tonal production is in parallel with the results of a study on tone perception development, in which the boundaries of categorical pitch contour perception sharpened at the age of 6 (F. Chen *et al.*, 2017).

In contrast to F0 cues, which were fully mastered at the age of 7–8, spectral cues exhibited a different developmental pattern. For the adults, spectral cues were sufficient to distinguish all four tones of Mandarin, replicating the findings of Ryant *et al.* (2014), which was based on a different corpus and recording set-up. However, such a high level of contrastivity of the adults' production was not reached in the children's production of spectral cues, even for the oldest children in our study (Age 10–11). Therefore, the integration of spectral cues in tonal production continues to develop throughout childhood and into adulthood, even after F0 cues are mastered.

For both F0 cues and spectral cues, between the age of 4–5 and 7–8 was found to be the critical period at which the steepest developmental curve was observed. In particular, for F0 cues, both the dispersion of the tonal clusters in MDS plots and

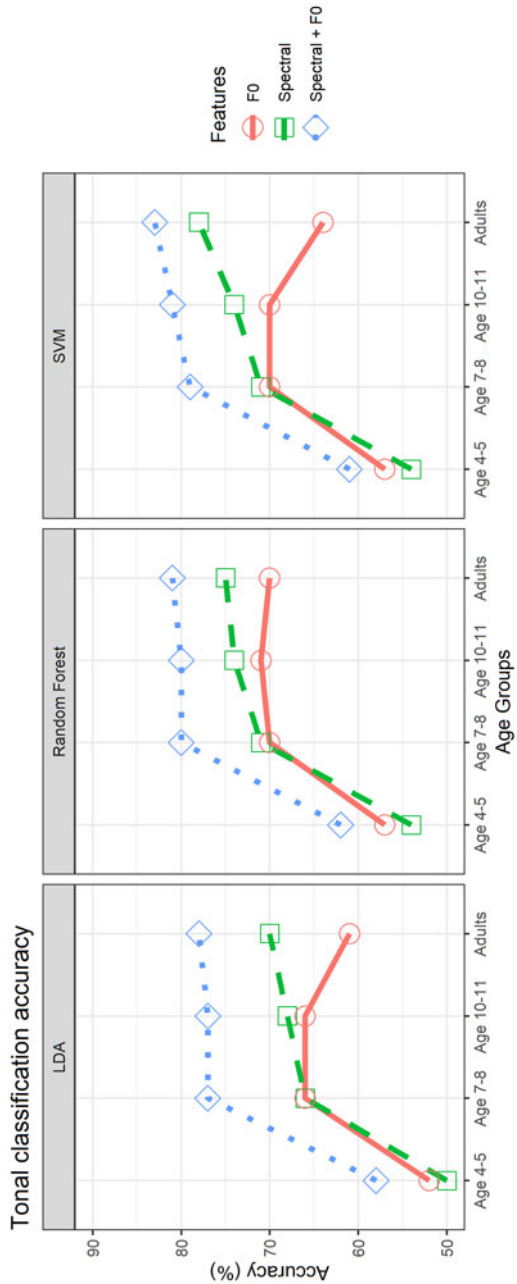


**Figure 2.** Tonal classification accuracy (chance level: 25%)

automatic classification results indicated that children of ages 7–8 and 10–11 not only reached but also exceeded adult-level tonal contrastivity: an overshoot in the production beyond the adult-level. A similar overshoot phenomenon has been observed in the development of the English vowel space area, where speakers aged 9 to 14 years had a larger vowel space area than adults (Pettinato *et al.*, 2016). While this overshoot phenomenon remains to be further explored, the delay in full development of spectral cues suggests the possibility that the hyperarticulation of F0 cues in tonal production of children may at least be partially explained as compensation for the underdeveloped allophonic cues.

As tone production and perception involve rich information from the voice source, it is important to investigate cues beyond F0. Voice quality cues in the spectrum naturally co-vary with F0, yet these physiologically correlated cues can also be linguistically coded (i.e., in the tonal system). In this study, we have shown that these spectral cues become increasingly useful in manifesting tone contrastivity. This developmental pattern can be explained in two ways: (i) children actively learn to integrate spectral cues in their production of tones as part of their linguistic development, or (ii) F0 and voice quality are physiologically correlated differently for children and adults, due to differences in the physiological control over their voices. Though we have not ruled out the second explanation, we argue that the first explanation is more plausible, given that adults are fully capable of making tonal contrasts even with reduced F0 cues, with the ability to incorporate other useful cues in tone production. Conversely, children, without fully-developed spectral cues, resort to producing greater F0 contrasts to achieve a comparable level of tonal contrastivity. Future research on the development of spectral cues in perception should further identify when and how children attune to these physiologically correlated cues in tone recognition.

## Conclusion

Using a semi-spontaneous speech corpus, we have presented evidence from multi-dimensional scaling and machine-learning classification of Mandarin tonal productions from children of different age groups (4- to 11-years old) and adults, suggesting that children's production of tones achieves and exceeds the adult level of tonal distinction at the age of 7 to 8 using F0 cues, exhibiting an overshoot in tone production with regard to F0. Both F0 and the allophonic spectral cues also exhibit a sharp growth spurt between ages 4 to 5 and 7 to 8; however, unlike the F0 cues, the adult-like use of spectral cues in encoding tonal contrasts is not achieved even at the age of 10 to 11. Adults, who are fully capable of producing the F0 patterns of the tones, have reduced F0 distinctions in tones, but possess sufficient and reliable spectral cues that can clearly distinguish all four tonal categories. This difference in the developmental curves of F0 and spectral cues in Mandarin tone production suggests that secondary cues may continue to develop even after the mastery of primary cues.

**Supplementary Material.** For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S0305000920000239

## References

**Astruc, L., Payne, E., Post, B., Vanrell, M. d. M., & Prieto, P.** (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, *56*(2), 229–253. doi:10.1177/0023830912460494

**Belotel-Grenié, A., & Grenié, M.** (1994). Phonation types analysis in Standard Chinese. In *The 3rd International Conference on Spoken Language Processing* (Vol. 94, pp. 343–346).

**Chen, A., Esteve-Gibert, N., Prieto, P., & Redford, M.** (unpublished observations). Development of phrasal prosody from infancy to late childhood. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Prosody*.

**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. doi:10.1017/S0305000916000581

**De Ruiter, L. E.** (2010). *Studies on intonation and information structure in child and adult German* (Doctoral dissertation, Radboud University Nijmegen Nijmegen).

**Galligan, R.** (1987). Intonation with single words: Purposive and grammatical use. *Journal of Child Language*, *14*(1), 1–21. doi:10.1017/S0305000900012708

**Hua, Z., & Dodd, B.** (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of child language*, *27*(1), 3–42. doi:10.1017/S030500099900402X

**Iseli, M., Shue, Y.-L., & Alwan, A.** (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, *121*(4), 2283–2295. doi:10.1121/1.2697522

**Kawahara, H., Masuda-Katsuse, I., & De Cheveigne, A.** (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency based F0 extraction: Possible role of a repetitive structure in sounds. *Speech communication*, *27*(3–4), 187–207. doi:10.1016/S0167–6393(98)00085-5

**Keating, P. A., & Esposito, C.** (2007). Linguistic voice quality. *UCLA Working Papers in Phonetics*, *105*(6), 85–91.

**Keating, P., Esposito, C., Garellek, M., Khan, S., & Kuang, J.** (2011). Phonation contrasts across languages. In *Proceedings of ICPhS XVII*.

**Kong, Y.-Y., & Zeng, F.-G.** (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America*, *120*(5), 2830–2840. doi:10.1121/1.2346009

**Kreiman, J., & Gerratt, B. R.** (2010). Perceptual sensitivity to first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America*, *128*(4), 2085–2089.

**Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N.** (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research*. doi:10.1044/1092-4388(2007/042)

**Kuang, J.** (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693–1706. doi:10.1121/1.5003649

**Kuang, J., & Liberman, M.** (2018). Integrating voice quality cues in the pitch perception of speech and non-speech utterances. *Frontiers in Psychology*, *9*, 2147.

**Lee, S., Potamianos, A., & Narayanan, S.** (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. doi:10.1121/1.426686

**Li, C. N., & Thompson, S. A.** (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185–199. doi:10.1017/S0305000900001598

**Liaw, A., & Wiener, M.** (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22. Retrieved from https://CRAN.R-project.org/doc/Rnews/

**Liu, P., Chen, Z., Larson, C. R., Huang, D., & Liu, H.** (2010). Auditory feedback control of voice fundamental frequency in school children. *The Journal of the Acoustical Society of America*, *128*(3), 1306–1312. doi:10.1121/1.3467773

**Liu, S., & Samuel, A. G.** (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*(2), 109–138. doi:10.1177/00238309040470020101

**Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F.** (2018). *E1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien*. R package version 1.7-0. Retrieved from https://CRAN.R-project.org/package=e1071

**Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G.L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H.** (2018). *Vegan: Community ecology package*. R package version 2.5-2. Retrieved from https://CRAN.R-project.org/package=vegan

**Patel, R., & Brayton, J. T.** (2009). Identifying prosodic contrasts in utterances produced by 4-, 7-, and 11-year-old children. *Journal of Speech, Language, and Hearing Research*, *52*(3), 790–801. doi:10.1044/1092-4388(2008/07-0137)

**Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V.** (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, *54*, 1–14. doi:10.1016/j.wocn.2015.07.002

**R Core Team**. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

**Ryant, N., Yuan, J., & Liberman, M.** (2014). Mandarin tone classification without pitch tracking. In *2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)* (pp. 4868–4872). IEEE.

**Shiller, D. M., Gracco, V. L., & Rvachew, S.** (2010). Auditory-motor learning during speech production in 9–11-year-old children. *PLoS ONE*, *5*(9), e12975. doi:10.1371/journal.pone.0012975

**Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K.** (2011). Voicesauce: A program for voice analysis. In *Proceedings of the ICPhS XVII*, *1846–1849*.

**Sundberg, J.** (1994). Vocal fold vibration patterns and phonatory modes. *STL-QPSR*, *35*, 69–80.

**Tingley, B. M., & Allen, G. D.** (1975). Development of speech timing control in children. *Child Development*, *46*(1), 186–194. doi:10.2307/1128847

**Venables, W. N., & Ripley, B. D.** (2002). *Modern applied statistics with S* (Fourth). New York: Springer. doi:10.1007/978-0-387-21706-2

**Whalen, D. H., & Xu, Y.** (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25–47. doi:10.1159/000261901

**Wong, P.** (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434–443. doi:10.1121/1.4768883

**Wong, P., & Strange, W.** (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PloS ONE*, *12*(8), e0182337. doi:10.1371/journal.pone.0182337

**Xu, Y.** (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*(1), 55–105.

**Yang, A., & Chen, A.** (2018). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, *38*(1), 26–46. doi:10.1177/0142723717733920

**Yeung, H. H., Chen, K. H., & Werker, J. F.** (2013). When does native language input affect phonetic perception? the precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139.

**Yuan, J., Ryant, N., & Liberman, M.** (2014). Automatic phonetic segmentation in Mandarin Chinese: Boundary models, glottal features and tone. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2539–2543). doi:10.1109/ICASSP.2014.6854058

**Cite this article:** Rhee N, Chen A, Kuang J (2020). Going beyond F0: The acquisition of Mandarin tones. *Journal of Child Language* 1–12. https://doi.org/10.1017/S0305000920000239