



BRIEF RESEARCH REPORT

Due to increased variability, the expanded vowel and tone space in Mandarin IDS does not lead to enhanced contrasts

Ping Tang¹ , Nan Xu Rattanasone² , Ivan Yuen³, Katherine Demuth² and Titia Benders⁴ 

¹School of Foreign Studies, Nanjing University of Science and Technology Jiangsu, Nanjing, Jiangsu, China,

²Department of Linguistics, The Centre for Language Sciences (CLaS), Macquarie University, North Ryde,

NSW Australia, ³Department of Language Science and Technology, Universität des Saarlandes, Saarbrücken,

Germany and ⁴Faculty of Humanities, University of Amsterdam, Amsterdam, Netherlands

Corresponding author: Tita Benders; Email: a.t.benders@uva.nl

(Received 05 May 2024; revised 29 January 2025; accepted 04 March 2025)

Abstract

Expanded vowel or tone space in IDS has traditionally been interpreted as evidence of enhanced acoustic contrasts. However, emerging evidence from various languages shows that the within-category acoustic *variability* of vowels and tones also increases in IDS, offsetting the benefit of space expansion and leading to non-enhanced, or reduced acoustic contrasts. This study re-analysed a corpus of Mandarin IDS and ADS, showing that, relative to ADS, vowels and tones in IDS display greater variability, resulting in non-enhanced contrasts. Thus, given increased variability, expanded vowel or tonal space in IDS may not necessarily equate to enhanced acoustic contrasts.

Keywords: infant-directed speech; Mandarin Chinese; vowels and tones

摘要

传统研究认为, 儿向语(Infant-Directed Speech, IDS)中元音或声调空间的扩展是声学对比度增强的证据。然而, 越来越多来自不同语言的研究表明, 儿向语中音位内部声学变异性同步增加, 这种现象抵消了空间扩展带来的语音强化效益, 最终导致声学对比度未增强甚至减弱。本研究通过对普通话儿向语与成人指向语(Adult-Directed Speech, ADS)语料库的再分析发现: 相较于成人指向语, 儿向语中的元音和声调表现出更大的声学变异性, 致使音位间对比度未能实现有效强化。因此, 在声学变异性增加的条件下, 儿向语中扩展的元音或声调空间并不必然等同于声学对比度的增强。

1. Introduction

Infants show remarkable abilities in acquiring complex language systems in a short amount of time. It has been suggested that early language input, which comes (partly) in the form of “infant-directed speech” (IDS), plays a critical role in early language

acquisition (Kuhl *et al.*, 1997). Many studies have demonstrated that, relative to adult-directed speech (ADS), phonological categories such as vowels and tones (in tonal languages) in IDS occupy expanded acoustic space, which has typically been interpreted as a sign of enhanced acoustic contrasts, potentially providing infants with acoustically more distinct phonological categories (Kuhl *et al.*, 1997; Liu, Kuhl, & Tsao, 2003; Uther, Knoll, & Burnham, 2007). However, accumulating evidence in recent years has shown that the (within-category) acoustic variability of vowels and tones also increases in IDS as compared to ADS (Cox *et al.*, 2023; Cristia & Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024; Wang *et al.*, 2021). Critically, the increased vowel and tonal variability leads to non-enhanced or even reduced vowel and tonal contrasts as quantified by both statistical centroids (i.e., means) and variances (Cox *et al.*, 2023; Cristia & Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024; Wang *et al.*, 2021). These findings start to challenge the traditional view that IDS exhibits enhanced acoustic contrasts, which supports phonetic learning. Moreover, they raise questions as to whether increased variability and the potential consequences for acoustic contrasts are observed for both vowels and tones in the same language. The present study therefore focused on Mandarin Chinese IDS, where both expanded vowel and tonal spaces have been observed (Tang *et al.*, 2017), to gain a more comprehensive view of acoustic contrasts in IDS at both segmental and suprasegmental levels.

Traditionally, the vowel space has been conceptualised as an area defined by the centroid means of three peripheral or point vowels (typically /a/, /i/, and /u/) in the first (F1) and second formants (F2) (e.g., Kuhl *et al.*, 1997). On average, the vowel space is expanded in IDS relative to ADS (Lovcevic *et al.*, 2024), as observed in many languages, including English, Japanese, Mandarin, Norwegian, Russian, and Swedish (Andruski, Kuhl, & Hayashi, 1999; Burnham, Kitamura, & Vollmer-Conna, 2002; Cox *et al.*, 2023; Kuhl *et al.*, 1997; Liu, Kuhl, & Tsao, 2003; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024; Tang *et al.*, 2017); however, note that such expansion is not consistently found in all studies and languages (e.g., Norwegian: Englund & Behne, 2005; Jamaican English: Wassink, Wright, & Franklin, 2007; English: Green *et al.*, 2010; Dutch: Benders, 2013; Cantonese: Xu Rattanasone, *et al.*, 2013; Danish: Cox *et al.*, 2023). In tonal languages such as Mandarin Chinese and Cantonese, where lexical tones with different pitch contours are used to differentiate word meanings, an expanded tonal space has also been observed in IDS, as reflected by acoustically more distinct pitch contour features (Tang *et al.*, 2017; Xu Rattanasone, *et al.*, 2013). These expanded acoustic spaces have been widely interpreted as a sign of enhanced phonemic contrasts, which might provide infants with acoustically more distinct categories, facilitating phonetic learning (e.g., Kuhl *et al.*, 1997; Uther, Knoll, & Burnham, 2007; Xu Rattanasone *et al.*, 2013). This account is further supported by evidence that mothers' vowel space area is associated with their infants' speech discrimination ability as well as vocabulary size (Kalashnikova & Burnham, 2018; Liu, Kuhl, & Tsao, 2003).

Vowel and tonal spaces are defined only in terms of category centroids without considering the degree of variability within each category. However, to define acoustic contrastiveness, it is necessary to consider both category means and variances (Miyazawa *et al.*, 2017). IDS has been reported to have acoustically more variable formants or pitch contours across vowels and/or tones as compared to ADS, resulting in an increase in the within-category variability of vowels and/or tones (Cox *et al.*, 2023; Cristia & Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024; Wang *et al.*, 2021; but see McClay *et al.*, 2021 for counter-evidence in ni-Vanuatu IDS). Such increased variability has been shown to result in increased proximity or even overlap between vowel categories, as evident in

English, Japanese, and Norwegian, outweighing the benefit of vowel space expansion and resulting in non-enhanced acoustic vowel contrasts (Cristia and Seidl, 2014; Miyazawa et al., 2017; Rosslund et al., 2022, 2024). For instance, Miyazawa et al. (2017) examined three acoustic characteristics of IDS in the three Japanese vowels /a/, /i/, and /u/: vowel space area, acoustic variability for each vowel (in F1 and F2 values), and acoustic contrasts between vowels. That study measured acoustic contrasts using Mahalanobis distance, capturing both between-vowel distinctiveness and within-vowel variability. Although Japanese IDS is characterised by an expanded vowel space area, the acoustic variability of all vowel categories is also increased. As a result, the acoustic contrasts (Mahalanobis distance) are not significantly different between IDS and ADS. Similarly, Wang et al. (2021) found that, relative to ADS, although the tonal space area is expanded in Cantonese IDS, the increased within-category variability of lexical tones leads to non-enhanced tonal contrasts (calculated using a method similar to Mahalanobis distance).

These results thus begin to challenge the assumed link between, on the one hand, enhanced acoustic contrasts and expanded vowel/tone space in IDS and, on the other hand, the role vowel/tone space plays in phonetic learning. For instance, increased vowel variability in caregivers' IDS is found to negatively correlate with several of their infants' language outcomes at 18 and 24 months (Rosslund et al., 2022). Computational modelling also showed that IDS does not yield more discriminable or more robust vowel categories as compared to ADS, primarily due to the increased acoustic variability (Kirchhoff & Schimmel, 2005; Ludusan, Mazuka, & Dupoux, 2021). However, evidence regarding the increased vowel variability in IDS and the potential consequences for vowel contrasts has only come from a few languages (Danish by Cox et al., 2023, English by Cristia & Seidl, 2014, Japanese by Miyazawa et al., 2017, and Norwegian by Rosslund et al., 2022, 2024), and tonal variability in IDS has only been examined in one language, i.e., Cantonese (Wang et al., 2021). No study has thus far examined both vowel and tonal variability in the same language to gain a more holistic view of acoustic contrasts in IDS at both segmental and suprasegmental levels.

The present study therefore focused on vowel and tonal variability in Mandarin IDS, which is characterised by expanded vowel and tonal spaces (Tang et al., 2017). We asked (1) whether IDS is characterised by an increased within-category acoustic variability of vowels and tones as compared to that of ADS, and (2) whether IDS differs from ADS in terms of acoustic contrasts of vowels and tones. Based on previous findings (e.g., Cox et al., 2023; Cristia & Seidl, 2014; Miyazawa et al., 2017; Rosslund et al., 2022, 2024; Wang et al., 2021), we predicted that both vowels and tones would demonstrate greater variability in IDS, with non-enhanced or reduced acoustic contrasts compared to ADS.

2. Methods

2.1. Corpus

This study re-analysed the data presented in Tang et al. (2017), which included target stimuli consisting of three peripheral vowels (/a/, /i/, and /u/) and three simple-contour tones (T1, T2, and T4), produced by 15 Mandarin-speaking mothers of 11–13-month-old infants (7 boys and 8 girls) in IDS and ADS. These mothers were born and raised in Mandarin-speaking families in Northern China (Beijing, Hebei Province, or northeastern China), and they spoke only Mandarin Chinese to their infants at home. Target stimuli were disyllabic nouns that could be elicited with toys, i.e., “**pa2** chong2 (爬虫)” *worm*, “**pi2** qiu2 (皮球)” *ball*, and “**pu2** ti2 (菩提)” *Bodhi* for target vowels, and “zhen1 **zhu1** (珍珠)”

pearl, “shan1 **zhu2** (山竹)” *mangosteen*, and “guang1 **zhu4** (光柱)” *light stick* for target tones. Target productions in IDS were elicited from mothers in a play session with their infants using these toys. Target productions in ADS were elicited in a conversation between the mothers and a native Mandarin speaker about the play sessions. Recordings were made in a sound-proof room while mothers wore a head-mounted condenser microphone (AKG C520) connected to a solid-state recorder (Marantz PMD661MKII) in a shoulder bag. All noise-free realisations of vowel targets were included in the vowel analysis. As target tones were the second syllable of the disyllabic target words, following a T1 syllable, tone sandhi effect on target tones should be minimised. Furthermore, to minimise anticipatory tonal-coarticulation effects from following tones (Xu, 1994), the tonal analysis was restricted to target words in utterance-final position, defined as being followed by a pause longer than 2 seconds. The included data consisted of 997 vowel tokens (IDS: 572 tokens; ADS: 425 tokens) and 455 tonal tokens (IDS: 255 tokens; ADS: 200 tokens)¹.

2.2. Acoustic parameter extraction

As in Tang *et al.* (2017), vowels were characterised by F1 and F2 values (in Bark), extracted from the middle portion (40%–60%) of each target vowel production. Tones were originally measured in semitones (reference: 50Hz) and characterised by two pitch-contour parameters: slope and curvature. These parameters effectively capture the global pitch movement of Mandarin tones, which are the most important correlates in tonal perception, especially for infants (Gauthier, Shi, & Xu, 2007). To obtain these two parameters, a second-order polynomial was fitted to the pitch contour of each target tone, and the linear and quadratic trends were extracted as the pitch contour slope and curvature values. A positive or negative slope indicates a rising or falling contour; a positive or negative curvature value reflects a concave- or convex-shaped contour (see Tang *et al.*, 2019 for a more detailed description).

2.3. Calculation of dependent variables

This study used ellipse area and Mahalanobis distance as two measures in the following analysis because the former reflects the degree of within-category variability, and the latter acoustic contrastiveness between categories.

The vowel or tonal variability was characterised as the ellipse area of each category based on F1 and F2 values (Hartman, Ratner, & Newman, 2017; Formula 1), or pitch slope and curvature features (Formula 2), where a larger ellipse area indicates acoustically more variable tokens within a category (note that only speakers contributing at least two tokens

¹ As suggested by an anonymous reviewer, we examined the syntactic structures of utterances containing the tonal target words in IDS and ADS. The results suggest that the target tones in the two registers appeared in structurally comparable utterances. In IDS, target tones were carried by words in isolation (52%), SVO sentences (16%), noun phrases (NPs; 15%), VO sentences (7%), and other structures (10%). Similarly, in ADS, target tones were carried by words in isolation (43%), SVO sentences (24%), NPs (11%), VO sentences (11%), and other structures (11%). Additionally, the words with target tones appeared in comparable sentence types between the two registers, with 95% of IDS utterances and 99% of ADS utterances being statements. These utterances were independently transcribed by three linguistics graduate students, with an inter-rater reliability of 93%.

per vowel or tone were included, as the calculation of SD values requires at least two data points):

$$\text{Vowel Ellipse Area} = \pi \times 2SD_{F1} \times 2SD_{F2} \quad (1)$$

$$\text{Tonal Ellipse Area} = \pi \times 2SD_{\text{pitch slope}} \times 2SD_{\text{pitch curvature}} \quad (2)$$

Following Miyazawa et al. (2017), the acoustic contrast between two phonological categories was indexed by the averaged Mahalanobis distance (Mahalanobis, 1936), which captures the difference between members of a pair relative to their summed variance. The Mahalanobis distance between two vowels (e.g., /a/ and /i/) was calculated as follows:

$$\text{Mahalanobis distance between /a/ and /i/} = \sqrt{\frac{2 * \left((\text{Mean } F1_{/a/} - \text{Mean } F1_{/i/})^2 + (\text{Mean } F2_{/a/} - \text{Mean } F2_{/i/})^2 \right)}{SD_{F1_{/a/}}^2 + SD_{F2_{/a/}}^2 + SD_{F1_{/i/}}^2 + SD_{F2_{/i/}}^2}} \quad (3)$$

Similarly, the Mahalanobis distance between two tones (e.g., T1 and T2) was calculated as follows:

$$\text{Mahalanobis distance between T1 and T2} = \sqrt{\frac{2 * \left((\text{Mean Slope}_{T1} - \text{Mean Slope}_{T2})^2 + (\text{Mean Curvature}_{T1} - \text{Mean Curvature}_{T2})^2 \right)}{SD_{\text{Slope}_{T1}}^2 + SD_{\text{Curvature}_{T1}}^2 + SD_{\text{Slope}_{T2}}^2 + SD_{\text{Curvature}_{T2}}^2}} \quad (4)$$

When the variance increases in the denominator, Mahalanobis distance will become smaller, suggesting reduced acoustic contrast between categories. Conversely, when the variance decreases, the Mahalanobis distance will become larger, suggesting increased acoustic contrasts. The obtained Mahalanobis distances of all vowel or tonal pairs were averaged for each speaker, and these averages were compared between IDS and ADS across speakers.

2.4. Statistical analysis

Group comparisons were conducted using linear mixed-effects models, as implemented in the R package “lme4” (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2022). Best models with the simplest random effects were obtained through a backward elimination approach, starting with a maximal model as justified by the design (Barr, Levy, Scheepers, & Tily, 2013). Then, we used the *anova* function to identify non-significant random slopes and excluded those from the models to keep models parsimonious (Bates, Mächler, Bolker, & Walker, 2015; R code of all final models is provided with the description of each model). The residual plot and quantile–quantile (Q–Q) plot for each model were generated using the *resid_panel* function from the “ggResidpanel” package and were then visually inspected to confirm that the model residuals followed a normal distribution (Goode & Rey, 2019). All *p*-values were obtained using the *anova* function of the package “lmerTest,” providing omnibus main effects and interactions

between all factors using *F*-statistics (Kuznetsova, Brockhoff, & Christensen, 2017). When a main effect or its interaction were significant ($p < 0.05$), Tukey-HSD post-hoc comparisons were conducted using the “lsmeans” package (Lenth, 2018). The effect size (partial eta squared for *F*-statistics and Cohen’s *d* for *t*-statistics) was calculated using the *eta_sq* function of the package “sjstats” (Lüdtke, 2018), and the *cohensD* function of the package “lsr” (Navarro, 2015). The confidence interval (CI) for partial eta square was calculated using the *ci.pvaf* function of the package “MBESS” (Kelley, 2023).

3. Results

3.1. Vowel and tonal variability

Figure 1 illustrates the acoustic distribution of all vowel and tonal tokens in IDS and ADS, and Figure 2 shows the acoustic variability (i.e., ellipse area) for each vowel or tonal category (also see Appendix 1 for a detailed description). Two separate linear mixed-effects models were conducted on vowel and tonal ellipse areas to investigate the register difference in variability. Both models included two fixed factors: “Register” (ADS and IDS; dummy coded with reference level: ADS) and either “Vowel” (/a/, /i/, and /u/; dummy coded with reference level: /a/) or “Tone” (T1, T2, and T4; dummy coded with reference level: T1), and a random intercept by “Speaker” (R code of the two final models:

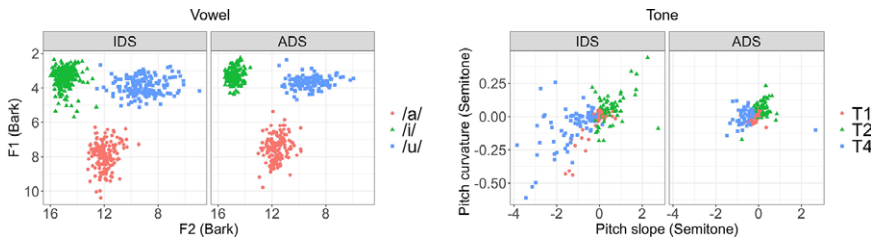


Figure 1. Acoustic distribution of all vowel and tonal tokens in IDS and ADS.

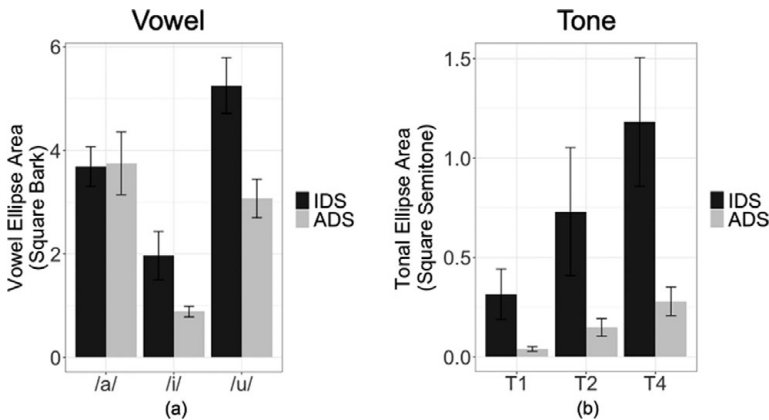


Figure 2. Ellipse area as an index of (a) vowel variability across vowels and registers or (b) tonal variability across tones and registers, with error bars indicating ± 1 standard error.

Vowel ellipse \sim Register * Vowel + (1|Speaker)²; Tonal ellipse \sim Register * Tone + (1|Speaker)³.

The results of the vowel variability model showed a significant interaction between “Register” and “Vowel” (Appendix 2). Tukey-HSD post-hoc tests showed that, relative to ADS, variability increased significantly in IDS for /i/ and /u/ (/i/: $\beta = 1.080$, SE = 0.47, $df = 39$, $t = 2.308$, $p = 0.026$, Cohen’s $d = 0.82$; /u/: $\beta = 2.180$, SE = 0.47, $df = 39$, $t = 4.660$, $p < 0.001^{***}$, Cohen’s $d = 1.22$), but there was no evidence for an increased variability of /a/ ($\beta = -0.063$, SE = 0.47, $df = 39$, $t = -0.134$, $p = 0.894$, Cohen’s $d = 0.03$).

The results of the tonal variability model showed a significant main effect of “Register” (Appendix 3), suggesting increased tonal variability in IDS as compared to ADS. Therefore, consistent with our hypotheses, IDS increased variability for the high vowels (see our explanation for the unexpected result for /a/ in the discussion) and across the investigated tones, providing evidence for increased within-category variability in (most) Mandarin Chinese vowels and tones.

3.2. Acoustic contrasts between vowels or tones

Figure 3 illustrates how acoustic contrasts (averaged Mahalanobis distance) differ between IDS and ADS, over (a) vowels or (b) tones (also see Appendix 4 for a detailed description). Two separate linear mixed-effect models were constructed using the averaged Mahalanobis distance for either vowels or tones as the dependent measure. Both models included one fixed factor “Register (ADS and IDS; dummy coded with reference level: ADS)” and a random intercept by “Speaker” (R code of the two final models: Averaged Mahalanobis distance for vowels \sim Register + (1 | Speaker)⁴; Averaged Mahalanobis distance for tones \sim Register + (1 | Speaker)⁵). The results of both models showed that the main effect of “Register” was not significant on averaged Mahalanobis distance for vowels ($F(1, 14) = 4.238$, $p = 0.059$, $\eta_p^2 = 0.23$ [95% CI: 0.00, 0.52]) or tones ($F(1, 27) = 0.002$, $p = 0.968$, $\eta_p^2 = 0.00006$ [95% CI: 0.00, 0.02]).

Null results for the “Register” effect on both vowels and tones provide no evidence for or against enhanced acoustic contrasts in IDS. However, if anything, the vowel results are compatible with larger contrasts in ADS, with the 95% CI ranging from $\eta_p^2 = 0.00$, suggesting no difference between the registers, to $\eta_p^2 = 0.56$, suggesting (much) larger contrasts in ADS. We therefore confidently interpret the vowel results as evidence that vowel contrasts are *not* enhanced in Mandarin IDS as compared to ADS. While the tonal data do not rule out enhancement in IDS, the point estimate of the associated effect size $\eta_p^2 = 0.00006$ is considered “small” in standard interpretations of effect size, and the upper bound of the CI $\eta_p^2 = 0.02$ falls well below a “medium” effect size of $\eta_p = 0.06$ (see Richardson, 2011 for a review). Therefore, we tentatively interpret these frequentist null results as evidence that tonal contrasts are not meaningfully enhanced in Mandarin IDS, as compared to ADS.

4. Discussion

The aim of the current study was to better understand the acoustic characteristics of IDS by investigating variability within vowels and tones in Mandarin IDS and its consequence

²Initial maximal model: Vowel ellipse \sim Register * Vowel + (1 + Register * Vowel | Speaker)

³Initial maximal model: Tonal ellipse \sim Register * Tone + (1 + Register * Tone | Speaker)

⁴Initial maximal model: Averaged Mahalanobis distance of vowels \sim Register + (1 + Register | Speaker)

⁵Initial maximal model: Averaged Mahalanobis distance of tones \sim Register + (1 + Register | Speaker)

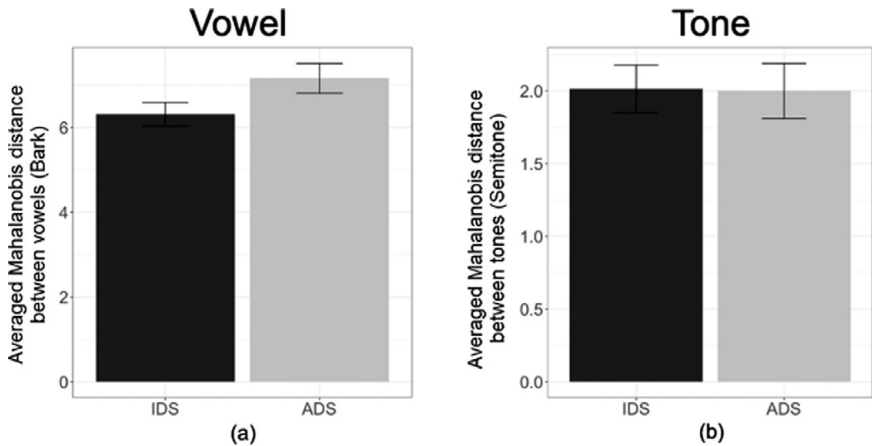


Figure 3. Averaged acoustic contrasts (Mahalanobis distance) for (a) vowels or (b) tones in IDS vs. ADS.

for acoustic contrasts. Reanalysing data from Tang *et al.* (2017), which showed expansion of both vowel and tonal spaces in IDS, the present study showed increased acoustic variability of both vowels (except for /a/, which will be discussed later) and tones. This, in turn, resulted in no evidence for enhanced contrasts between vowels or tones in IDS compared to ADS.

Our findings of the increased vowel and tonal variability in Mandarin IDS are consistent with the increased vowel variability reported for IDS in non-tonal languages such as Danish, English, Japanese, and Norwegian (Cox *et al.*, 2023; Cristia & Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024), and the increased tonal variability reported for Cantonese IDS (Wang *et al.*, 2021). This suggests that an increase in variability might be a common feature of IDS vowels and tones across languages (albeit, see McClay *et al.* 2021, for non-increased variability in ni-Vanuatu IDS).

Furthermore, the non-enhanced vowel and tonal contrasts in Mandarin IDS also confirm and extend previous findings on IDS in the aforementioned non-tonal languages (e.g., Cox *et al.*, 2023; Cristia & Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024), challenging the traditional view that IDS might provide infants with acoustically more distinct phonetic targets for the benefit of language development. The current study thus underscores the necessity and importance of characterising acoustic contrasts beyond statistical centroids (e.g., means) by including variance. This is critical to our understanding of the nature of IDS (Cox *et al.*, 2023; Cristia & Seidl, 2014; Hartman, Ratner, & Newman, 2017; McMurray *et al.*, 2013; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024; Wang *et al.*, 2021), as well as its potential influences on phonetic learning (will be discussed later; Eaves *et al.*, 2016; Kirchhoff & Schimmel, 2005; Ludusan *et al.*, 2021; Rosslund *et al.*, 2022). Compared to the commonly used acoustic space, which only considers the centroid means of categories, the Mahalanobis distance offers an approach to measuring acoustic contrasts that accounts for both between and within-category distances. Although the current IDS data has previously shown expanded vowel and tone spaces (Tang *et al.*, 2017), the Mahalanobis distance analyses here suggest that expanded spaces do not necessarily imply enhanced contrasts, aligning with similar findings from English, Japanese, and Norwegian (Cristia and Seidl, 2014; Miyazawa *et al.*, 2017; Rosslund *et al.*, 2022, 2024). A next step in this line of enquiry would be to examine

whether Mahalanobis distance is related to perception and learning. Considering that infants' language outcomes are positively associated with vowel space enhancement, defined by just the centroids (Kalashnikova & Burnham, 2018; Liu, Kuhl, & Tsao, 2003), and negatively associated with vowel variability (Rosslund et al., 2022), such work would further clarify the relationship between various dimensions of acoustic contrast and infants' actual language acquisition process.

It should be noted that there is no evidence for (or against) the variability of /a/ differing between IDS and ADS. This might be related to the fact that this vowel was relatively variable even in ADS (Figure 1). Previous studies have demonstrated that low vowels (such as /a/) usually display larger inter- and intra-speaker variability than non-low or high vowels (such as /i/ and /u/), due to the lack of lingual bracing against the palate during the articulation of low vowels (Gick et al., 2017; Whalen et al., 2018). The observed vowel specificity in variability calls for future studies of IDS to include more vowel types in order to obtain a full picture of the variability that infants are exposed to across the vowel system.

Our results raise questions about possible sources of increased vowel and tonal variability in IDS. One possibility is that increased variability could be a by-product of vowel and/or tonal space expansion, i.e., the expanded space allows for mothers producing more variable instances within each category (Kuhl et al., 1997). However, a series of exploratory analyses of the current data did not provide strong evidence for associations between vowel or tonal space area and acoustic variability (as indexed by ellipse area) in IDS (Appendix 5). A second option is that the increased vowel and tonal variability might follow from more variable prosodic patterns in IDS, which have been suggested to serve an attention-attracting function (Fernald, 1989; Fernald & Kuhl, 1987). This possibility is supported by some classic phonetic studies showing that prosody can impact phonetic realisation (Gay, 1978; Lindblom & Sundberg, 1971; Sapir, 1989). More work is thus needed to further examine potential prosodic effects on increased vowel and tonal variability in IDS.

Our study also raises the question of whether infants can benefit from IDS during language learning, given that vowels and tones do not become (much) more distinct once the increased variability is considered. On the one hand, increased within-category variability can be beneficial for infants in learning words and developing robust phonological representations (Singh, 2008; Thiessen, 2011). For example, Singh (2008) found that 7.5-month-old infants trained with high-variability words (produced by the same speaker in different emotional states) are better at recognizing trained words and rejecting similar-sounding words than those who were trained with low-variability words (produced by the same speaker in a single emotional state). Thiessen (2011) also showed that 15–16-month-old infants exposed to the /d/-/t/ contrast with large variability, i.e., produced in multiple word contexts, were able to discriminate this contrast in novel word contexts, while those exposed to this contrast in a single word context failed to do so. On the other hand, recent evidence has shown that vowel variability in IDS is negatively correlated with expressive vocabulary size at 18 months (in Norwegian, Rosslund et al., 2022) and various language abilities at 24 months (in American English; Hartman, Ratner, & Newman, 2017). Therefore, it remains an open question whether the benefits of increased variability as observed in (experimental) word-learning contexts extend to multiple aspects of naturalistic language acquisition, and under what conditions variability may instead hinder language development.

There are several limitations to this study. First, while this study has partially controlled for the phonological environment of target vowels and tones by restricting

them to the first (vowels) or second (tones) syllable of disyllabic nouns, there will be factors that could influence the acoustic realisation of vowels and tones in IDS, such as the pragmatic context (Katz, Cohn & Moore, 1996; Neer *et al.*, 2024), infants' responses during interaction (Outters *et al.*, 2020; Smith & Trainor, 2008), and parents' emotional state during speech productions (Benders, 2013; Kaplan *et al.*, 2002; Lam-Cassettari & Kohlhoff, 2020). The goal of the present study was to make a broad comparison between IDS and ADS, including all the inherent differences between the registers, thereby directly extending previous IDS findings from other languages (e.g., Cantonese, English, Japanese, and Norwegian) to Mandarin. Future studies could examine whether these and other factors contribute to the increased acoustic variability in IDS, with more controlled designs and larger datasets to pinpoint sources of within-speaker variance and their potential implications for phonetic learning.

Second, by only analysing tones in the second syllable of a disyllabic word in the utterance-final position, the current tonal results cannot be generalised to other positions. Our choice is motivated by the need to control for the carryover tonal coarticulation effect from the first syllable (T1) of the disyllabic word while minimizing the effect of anticipatory tonal coarticulation from the following words. Future research could examine tonal variability in a wider range of contexts to provide a more comprehensive understanding of how tones are realised in IDS.

5. Conclusions

The current study demonstrated that Mandarin vowels and tones in IDS exhibit increased (within-category) variability, resulting in non-enhanced vowel and tonal contrasts. Thus, IDS might not be able to provide infants with acoustically well-specified phonetic targets that facilitate phonological acquisition as directly as often assumed. Rather, IDS might provide infants with considerable within-speaker phonetic variability. Future research should examine the linguistic and communicative factors that drive the increased acoustic variability in IDS, as well as its potential positive or negative impacts on language acquisition.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0305000925000133>.

Acknowledgements. The research is funded by the National Social Science Fund of China (20CYY012).

References

- Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *The Journal of the Acoustical Society of America*, **105**(2_Supplement), 1095–1096.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* **67**(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benders, T. (2013). Mommy is only happy! Dutch mothers' realization of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development* **36**(4), 847–862. <https://doi.org/10.1016/j.infbeh.2013.09.001>
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science* **296**(5572), 1435. <https://doi.org/10.1126/science.1069587>
- Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour* **7**(1), 114–133. <https://doi.org/10.1038/s41562-022-01452-1>

- Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language* 41(4), 913–934. <https://doi.org/10.1017/S0305000912000669>
- Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review* 123(6), 758–771. <https://doi.org/10.1037/rev0000031>
- Englund, K. T., & Behne, D. M. (2005). Infant-directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research* 34(3), 259–280. <https://doi.org/10.1007/s10936-005-3640-7>
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development* 10(3), 279–293. [https://doi.org/10.1016/0163-6383\(87\)90017-8](https://doi.org/10.1016/0163-6383(87)90017-8)
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16(3), 477–501. <https://doi.org/10.1017/S0305000900010679>
- Gauthier, B., Shi, R., & Xu, Y. (2007). Learning phonetic categories by tracking movements. *Cognition* 103(1), 80–106. <https://doi.org/10.1016/j.cognition.2006.03.002>
- Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America* 63(1), 223–230. <https://doi.org/10.1121/1.381717>
- Gick, B., Allen, B., Roewer-Després, F., & Stavness, I. (2017). Speaking tongues are actively braced. *Journal of Speech, Language, and Hearing Research* 60(3), 494–506. https://doi.org/10.1044/2016_JSLHR-S-15-0141
- Goode, K., & Rey, K. (2019). ggResidpanel: Panels and interactive versions of diagnostic plots using 'ggplot2'. *R package version 0.3.0*. <https://cran.r-project.org/web/packages/ggResidpanel/index.html>
- Green, J., Nip, I., Wilson, E., Mefferd, A., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research* 53(6) 1529–1542. [https://doi.org/10.1044/1092-4388\(2010/09-0005\)](https://doi.org/10.1044/1092-4388(2010/09-0005))
- Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language* 44(5), 1140. <https://doi.org/10.1017/S030500916000520>
- Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language* 45(5), 1035–1053. <https://doi.org/10.1017/S0305000917000629>
- Kaplan, P. S., Bachorowski, J. A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science* 13(3), 268–271. <https://doi.org/10.1111/1467-9280.00449>
- Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, 67(1), 205–217.
- Kelley, K. (2023). *MBESS: The MBESS R package* (Version 4.9.3). <https://CRAN.R-project.org/package=MBESS>
- Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *The Journal of the Acoustical Society of America* 117(4), 2238–2246. <https://doi.org/10.1121/1.1869172>
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. L., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277(5326), 684–686. <https://doi.org/10.1126/science.277.5326.684>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lam-Cassettari, C., & Kohlhoff, J. (2020). Effect of maternal depression on infant-directed speech to prelinguistic infants: Implications for language development. *PLOS ONE* 15(7), e0236787. <https://doi.org/10.1371/journal.pone.0236787>
- Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician* 34(4), 216–221. <https://doi.org/10.1080/00031305.1980.10483093>
- Lovcic, I., Benders, T., Tsuji, S., & Fusaroli, R. (2024). Vowel-hyperarticulation in infant-directed speech: A multi-method meta-analysis. *PsyArXiv*. https://osf.io/preprints/psyarxiv/2sfye_v1
- Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America* 50(4B), 1166–1179. <https://doi.org/10.1121/1.1912750>

- Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* 6(3), F1–F10. <https://doi.org/10.1111/1467-7687.00275>
- Lüdtke, D. (2018). sjstats: Statistical functions for regression models (Version 0.14.0) [R package]. <https://cran.r-project.org/web/packages/sjstats/index.html>
- Ludusan, B., Mazuka, R., & Dupoux, E. (2021). Does infant-directed speech help phonetic learning? A machine learning investigation. *Cognitive Science* 45(5), e12946. <https://doi.org/10.1111/cogs.12946>
- Mahalanobis, P. C. (1936). *On the generalized distance in statistics*. National Institute of Science of India.
- McClay, E. K., Cebiglu, S., Broesch, T., & Yeung, H. H. (2021). Rethinking the phonetics of baby-talk: Differences across Canada and Vanuatu in the articulation of mothers' speech to infants. *Developmental Science*, e13180. <https://doi.org/10.1111/desc.13180>
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant-directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition* 129(2), 362–378. <https://doi.org/10.1016/j.cognition.2013.07.015>
- Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition* 166, 84–93. <https://doi.org/10.1016/j.cognition.2017.05.003>
- Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.6) [R package]. University of New South Wales. <https://cran.r-project.org/web/packages/learningstatistics/>
- Neer, E., Brahmabhatt, A., Walsh, C., & Warlaumont, A. S. (2024). Pitch Characteristics of Real-World Infant-Directed Speech Vary with Pragmatic Context, Perceived Adult Gender, and Infant Gender. PsyArXiv. <https://doi.org/10.31234/osf.io/2z8mf>
- Outters, V., Schreiner, M. S., Behne, T., & Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy* 25(4), 478–499. <https://doi.org/10.1111/infa.12334>
- R Core Team. (2022). R: A language and environment for statistical computing [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6(2), 135–147. <https://doi.org/10.1016/j.edurev.2010.12.001>
- Rosslund, A., Mayor, J., Óturai, G., & Kartushina, N. (2022). Parents' hyper-pitch and low vowel category variability in infant-directed speech are associated with 18-month-old toddlers' expressive vocabulary. *Language Development Research* 2(1), 223–267. <https://doi.org/10.34842/2022.0547>
- Rosslund, A., Mayor, J., Mundry, R., Singh, A. P., Cristia, A., & Kartushina, N. (2024). A longitudinal investigation of the acoustic properties of infant-directed speech from 6 to 18 months. *Royal Society Open Science* 11(11), 240572. <https://doi.org/10.1098/rsos.240572>
- Sapir, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. *Journal of Voice* 3(1), 44–51. [https://doi.org/10.1016/S0892-1997\(89\)80121-3](https://doi.org/10.1016/S0892-1997(89)80121-3)
- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, 106(2), 833–870.
- Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy* 13(4), 410–420. <https://doi.org/10.1080/15250000802188719>
- Thiessen, E. D. (2011). When variability matters more than meaning: the effect of lexical forms on use of phonemic contrasts. *Developmental Psychology*, 47(5), 1448.
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America* 142(2), 493–503. <https://doi.org/10.1121/1.4995998>
- Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 62(5), 1309–1325.
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49(1), 2–7. <https://doi.org/10.1016/j.specom.2006.10.003>
- Wang, L., Kalashnikova, M., Kager, R., Lai, R., & Wong, P. (2021). Lexical and prosodic pitch modifications in Cantonese infant-directed speech. *Journal of Child Language* 1–27. <https://doi.org/10.1017/S0305009200000707>

- Wassink, A. B., Wright, R. A., & Franklin, A. D.** (2007). Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35(3), 363–379. <https://doi.org/10.1016/j.wocn.2006.07.002>
- Whalen, D. H., Chen, W. R., Tiede, M. K., & Nam, H.** (2018). Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics* 68, 1–14. <https://doi.org/10.1016/j.wocn.2018.01.003>
- Xu Rattanasone, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U.** (2013). Vowel hyperarticulation in parrot-, dog-, and infant-directed speech. *Anthrozoös* 26(3), 373–380. <https://doi.org/10.2752/175303713X13697429463592>
- Xu, Y.** (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America* 95(4), 2240–2253. <https://doi.org/10.1121/1.408684>

Cite this article: Tang, P., Xu Rattanasone, N., Yuen, I., Demuth, K., & Benders, T. (2025). Due to increased variability, the expanded vowel and tone space in Mandarin IDS does not lead to enhanced contrasts. *Journal of Child Language* 1–13, <https://doi.org/10.1017/S0305000925000133>