Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Faster and Smoother: Fluency in Chinese Child-directed Speech

Mengru Han[1], *Lianghui Yang*[1], *Yan Gu*[2,3]

[1]Department of Chinese Language and Literature, East China Normal University, China
[2]Department of Psychology, University of Essex, UK
[3]Department of Experimental Psychology, University College London, UK
mrhan@zhwx.ecnu.edu.cn, 10213330408@stu.ecnu.edu.cn, yan.gu@essex.ac.uk

Abstract

Child-directed speech (CDS) is often believed to have a slower speaking rate than adult-directed speech (ADS). This study examined the fluency between CDS and ADS as well as the individual differences in mothers' speaking rates. We annotated 2917 utterances in a corpus of Chinese ADS and CDS, where 19 mothers told the same story to their 24-month-old children and an adult. We coded and compared the fluency measures between ADS and CDS: speech rate (SR, including utterance-internal pauses), articulation rate (AR, excluding utterance-internal pauses), frequencies of silent pauses, filled pauses, repairs, and repetitions. We have three main findings: (1) CDS was generally more fluent than ADS, with fewer silent and filled pauses. (2) Contrary to common belief, only 7 out of the 19 participants showed a decreased SR and AR in CDS. (3) There were no significant differences in SR or AR between CDS and ADS when the utterance length was shorter than 4 syllables, whereas CDS was significantly faster than ADS when utterances were longer than 5 syllables. This suggests that Chinese CDS is not slower but instead *faster* than ADS. These findings highlight language-specific and individual variations in the temporal aspects of CDS.

Index Terms: child-directed speech, fluency, speech rate, Chinese

1. Introduction

When talking to young children, caregivers use a unique speaking style called child-directed speech (CDS). Typically, CDS is characterized by a slower speaking rate compared to adult-directed speech (ADS), often seen as evidence for mothers to adjust their prosody to aid early language learning [1], [2]. However, this pattern may not hold true across all languages, age groups, or throughout the entire utterance (e.g., [3], [4]). Recent studies on Chinese, a syllable-timed language, have shown no evidence of slowing down in CDS compared to ADS[5]–[7]. Given that utterances in CDS are typically shorter than those in ADS, and utterance length can influence speaking rate [3], [8], it remains unclear whether Chinese CDS is slower than ADS when controlling for utterance length.

Speaking rate is an important component of speech fluency, alongside pauses, repetitions, and repairs [9]. Despite previous studies highlighting the effect of speaking rate or pausing on children's word learning as separate cues (e.g., [10]–[12]), there has been no systematic comparison of these fluency measures between ADS and CDS.

In addition, fluency can reflect individual speaking styles. For example, correlations have been found in fluency measures between a speaker's first (L1) and second (L2) languages [13], [14], as well as between different speech registers within the same language (e.g., child-directed vs. adult-directed broadcasting [15]). Despite CDS's adaptability to children's age and linguistic ability [16], [17], no research has explored whether CDS reflects an individual's speaking style in ADS.

This paper systematically compared fluency in Chinese CDS and ADS and examined the speaking rate as a function of utterance length, while also highlighting individual differences in CDS and ADS. The findings will reveal insights into the language-specificity of CDS, the temporal aspects of speech adjustments for listeners, and their possible impact on language learning.

1.1. Temporal modification in CDS and language learning

CDS has a slower speaking rate compared to ADS across languages such as German, English, Japanese, and Swedish [3], [16], [18]–[21]. These studies suggest that mothers may slow down their speech in CDS to support children's language learning. Slow speech offers several benefits for language learning. First, it aligns with children's developing brain. Nencheva and Lew-Williams [1, pp. 4] suggest that "Because infants have increased power in slower neural oscillations and a slower theta rhythm, IDS (i.e., infant-directed speech, author's note) *should* have a slower rhythm, otherwise infants' attention and processing would be frequently misaligned." Second, slow speech tends to be clearer, which can improve speech perception and language comprehension [22], [23]. Studies have demonstrated that children exhibit significantly improved word recognition with slower speech compared to faster speech in laboratory settings [24], [25]. Additionally, mothers' speaking rate in CDS has been found to predict children's vocabulary size [6], [19].

However, recent evidence challenges the view that mothers consistently speak more slowly in CDS. For example, in Japanese CDS, overall utterances are slower than in ADS, primarily due to phrase-final lengthening rather than a consistent slowing down throughout the entire utterance [3]. Dutch CDS is slower than ADS, particularly when introducing unfamiliar words, but no similar slowing down is observed in Chinese CDS [6]. Moreover, in the Chinese broadcasting context, CDS is not slower than ADS [5].

Previous studies on CDS speaking rate have raised several concerns. First, most studies measure the global speaking rate or pauses between utterances (as discussed in [3]; also see [26]), rather than focusing on the utterance level, where factors like utterance length can affect speech rate: articulation rate tends to be faster with longer utterances [8]. Second, the speech context in CDS and ADS often differs significantly. CDS typically includes natural or semi-natural interactions between mothers and children in laboratory or home settings, while ADS often results from casual conversations between caregivers and experimenters. These differing content and contexts may amplify distinctions between the two speech registers. Third,

10.21437/SpeechProsody.2024-2

researchers often conduct group comparisons instead of examining individual differences, leaving it unclear whether every mother slows down when talking to children.

1.2. Fluency in CDS and its impact on language learning

Speaking rate, defined as the "rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention under the temporal constraints of on-line processing." [20, pp. 26] (refer to [28] for a review of different definitions), is an important component of speech fluency. In addition to speech rate (including silent pauses) and articulation rate (excluding silent pauses), fluency measures often include pauses, repetitions, and repairs [9]. In general, being fluent is characterized by a faster speaking rate, fewer and shorter pauses, and fewer repetitions and repairs [28], though Bosker *et al*. [9] found that articulation rate and pausing are more important than repairs in listeners' perceived fluency. Research also shows that L2 fluency is related to L1 fluency, indicating speakers' individual speaking tyle [21], [26], [27].

While speech fluency has been the main area of interest in L2 research, it is typically ignored in the line of research that compares almost all aspects between ADS and CDS. It is often assumed that CDS is and should be more fluent than ADS. For example, Soderstrom and Morgan [11] suggest that "maternal speech may be highly well-formed because it comes packaged in short utterances that leave little opportunity for disfluency to manifest." It is important to note that disfluency is not exclusive to L2 production. Even when speaking L1, speakers do not always speak fluently [29]. When addressing young children, mothers can also have difficulties conceptualizing or processing, resulting in disfluency in their speech production [30]. So far, very few studies have compared the fluency between CDS and ADS, and they have rarely measured utterance-internal pauses. Bellinghausen *et al*. [31] compared the disfluency of nine German mothers' ADS and CDS and found that German ADS had more silent and filled pauses than CDS, but they did not measure speaking rate.

While pausing may be associated with speech disfluency, it can also benefit language learning. Research has shown that children can discriminate between fluent and disfluent speech as early as 22 months [11]. Children aged 28–32 months can reliably use the filled pauses "uh" and "um" to predict infrequent or discourse-new words [12]. In addition, preschoolers' word learning is facilitated by pausing in shared-book reading [10].

In short, there appears to be a contradiction between being fluent and slow at the same time in CDS, as fluency typically indicates speaking quickly, while CDS is associated with speaking slowly. It is unclear what role utterance length plays or how fluency measures such as pausing, repetitions, and repairs contribute to fluency in CDS compared to ADS.

1.3. Current study

This study aims to better understand the fluency differences between ADS and CDS using a fully-transcribed corpus with similar content in both speech registers. Specifically, we will address the following research questions:

First, is Chinese CDS faster than ADS at the utterance level? Although Han *et al*. [6], [7] did not find any significant differences in articulation rate between Chinese ADS and CDS at the global level, nor when mothers introduced familiar or unfamiliar words, we still predict that a more detailed measure would reveal that Chinese CDS is slower than ADS, in line with findings in other languages.

Second, is Chinese CDS more fluent than ADS? By examining fluency measures such as pausing, repetition, and repairs, we predict that CDS is more fluent than ADS, with fewer and shorter pauses, fewer repetitions, and repairs.

Additionally, what are the individual differences in mothers' temporal modifications in CDS? We examined them from two perspectives: (1) As speakers' fluency can be linked between their different languages (e.g., [14]) and different speech registers [15], we expect correlations in mothers' fluency between ADS and CDS. (2) Do all mothers speak more slowly in CDS? If there are individual differences in CDS, then not every mother would slow down their speaking rate in CDS.

2. Method

2.1. Speech corpus and participants

We conducted fluency measurements at the utterance level in a corpus of Chinese ADS and CDS [32]. The corpus contained speech from 19 Chinese-speaking mother-child dyads, with the children being 24 months old (10 girls and 9 boys; mean age = 24;13, age range = 23;27–24;30). All children were typically developing and had no reported language impairments or hearing problems.

The ADS and CDS were semi-spontaneous speech elicited in a shared-book reading task. The book consisted of 12 pages, each featuring a word on the left side and an illustration of the word on the right side. Mothers were asked to tell the story twice, once in ADS and once in CDS. To elicit CDS, the mothers were instructed to tell the story to their child as they normally would at home. To elicit ADS, mothers told the same story to an adult (a female Mandarin speaker) while the child was not present. The mothers were free to construct the story, except that they included the words given on each page. The order of the two speech registers was counterbalanced across participants. The fact that the mothers told the same story in both ADS and CDS allowed us to compare their speech modulation when the content was similar.

2.2. Data coding

2.2.1. Speech transcription and annotation

First, we used an automatic Chinese speech recognition tool developed by Iflytek (https://www.iflyrec.com) to transcribe speech data. These transcriptions were then segmented into utterances by two Chinese speakers. Following Martin *et al*. [3], utterances are defined as "[…] any pause longer than 200 ms which is preceded by an intonational phrase boundary (pauses not accompanied by an IP boundary were considered utterance-internal)". The corpus had a total of 3257 utterances, of which 2917 were produced by mothers. The transcriptions included utterance-internal silent pauses (threshold: 200 ms), filled pauses (e.g., *uh*, *um*, *zhege*, *nage*), repairs, and repetitions.

Second, all utterances were manually aligned with the speech in Praat [33]. The silent pauses were manually annotated. A trained native speaker (the first author) listened to each utterance to verify the accuracy of the transcription and the utterance boundaries. The intercoder reliability was 92%.

Third, a Python script was used to count the number of (phonological) syllables (utterance length), utterance-internal silent pauses, filled pauses, repairs, and repetitions for each

utterance. A Praat script was used to extract the duration of each utterance and the utterance-internal silent pauses. Based on these measurements, we calculated the fluency measures.

2.2.2. Fluency measurements

We obtained the following fluency measures at the utterance level: speech rate (SR, including utterance-internal pauses), articulation rate (AR, excluding utterance-internal pauses), frequencies of silent pauses, filled pauses, repairs, and repetitions (instances per 100 syllables).

2.3. Data analysis

In R [34], we first performed paired-sample t-tests to determine, at the group level, whether the fluency measures differed between ADS and CDS. For speech rate and articulation rate, as utterance length may influence speaking rate [8], we further examined whether utterance length (number of syllables in each utterance) interacted with speaking rate. For this analysis, we used a linear mixed-effects model implemented through the 'lmer' function from the 'lme4' package [35]. Both speech rate and articulation rate were log-transformed before analyses. The fixed effects were Condition (ADS/CDS) and utterance length, as well as their interaction. The random intercepts and slopes for Condition varied by Participant.

To examine individual differences, we compared the mean speech rate and articulation rate for each participant between the ADS and CDS conditions. We also conducted the Pearson's correlation test for the measures between them.

3. Results

3.1. Comparing fluency measures between ADS and CDS

Table 1 presents the means of the measures and demonstrates that both the articulation rate (p = 0.017) and speech rate (p = 0.009) were, unexpectedly, *faster* in CDS than in ADS. Additionally, there were significantly fewer silent pauses (p = 0.041) and filled pauses (p = 0.037), as well as shorter mean pause durations (p = 0.011) in CDS compared to ADS. However, the ratio of repetitions and repairs did not differ between the two conditions (p's > 0.1). These results suggest that Chinese CDS is not slower but instead faster and more fluent than ADS.

3.2. Effect of utterance length on speech rate/articulation rate in CDS and ADS

As shown in Table 1, the mean utterance length was shorter in CDS than in ADS. For speech rate, the main effect of condition (β = 0.10, 95% CI [0.05, 0.16], p = 0.001) and utterance length (β = 0.03, 95% CI [0.02, 0.03], p < 0.001) were both significant. Moreover, there was a significant interaction between utterance length and Condition (β = 0.01, 95% CI [0.01, 0.02], p < 0.001) (see Figure 1). This indicats that the differences between conditions are significantly different for longer utterances compared to shorter utterances.

When further probing the interaction using the emmeans() function in R to examine the threshold of utterance length, it was found that there were no significant differences in speech rate or articulation rate between CDS and ADS when the utterance length was 4 syllables or less (N = 939, p's > 0.05). However, CDS was significantly faster than ADS for utterances longer than 5 syllables (N = 1978, p's < 0.05). A similar pattern was observed for articulation rate, with a significant interaction between utterance length and condition (β = 0.01, 95% CI [0.004, 0.016], p = 0.001). Thus, the differences in speech rate and articulation rate between CDS and ADS were influenced by utterance length: with longer utterances in CDS being even faster than those in ADS.

Table 1: *Means and SDs of the fluency measures*

Measures	***Mean (SD)***		***T***	***r***
	ADS	CDS		
Articulation rate (*syllables/s*)	4.60 (0.48)	4.95 (0.46)	-2.64*	0.22
Speech rate (*syllables/s*)	4.49 (0.47)	4.89 (0.46)	-2.94**	0.18
Silent pause ratio (*n/100 syllables*)	2.14 (1.31)	1.42 (1.09)	2.20*	0.31
Filled pause ratio (*n/100 syllables*)	0.48 (0.83)	0.08 (0.14)	2.26*	0.51*
Repetition ratio (*n/100 syllables*)	0.12 (0.23)	0.23 (0.26)	-1.37	-0.03
Repair ratio (*n/100 syllables*)	0.40 (0.40)	0.29 (0.26)	0.96	0.004
Mean pause duration (*s*)	0.50 (0.24)	0.37 (0.09)	2.86*	0.50*
Utterance length (*n of syllables*)	8.30 (1.46)	6.63 (0.96)	5.91***	0.55*
Mean utterance duration (*s*)	1.89 (0.28)	1.42 (0.18)	8.44***	0.51*

*Notes: (1) Positive T-value means that ADS > CDS, negative T-value means ADS < CDS; (2) *p < .05; **p < .01; ***p < .001.*

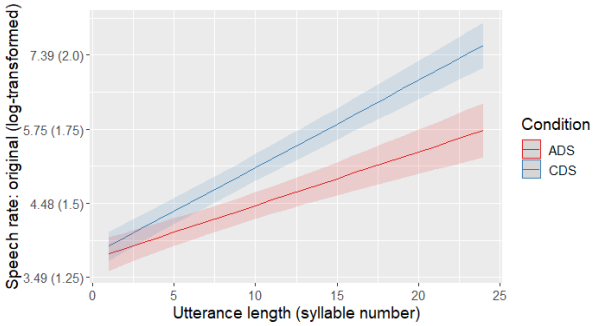

Figure 1: *Predicted relationship between utterance length and speech rate for each Condition (ADS/CDS). A similar pattern was observed for articulation rate.*

3.3. Individual differences in CDS temporal modifications

There were no significant correlations between ADS and CDS in most of the fluency measures (Table 1, p's > 0.1), except for the filled pause ratio and mean pause duration. This suggests that mothers' speaking style was only characterized by pausing behavior rather than speaking rate. Mothers who used more filled pauses and longer pause durations in ADS also used more filled pauses and longer pause durations in CDS. Furthermore, as shown in Figure 2, only 7 out of the 19 mothers exhibited a decreased speech rate and articulation rate in CDS.

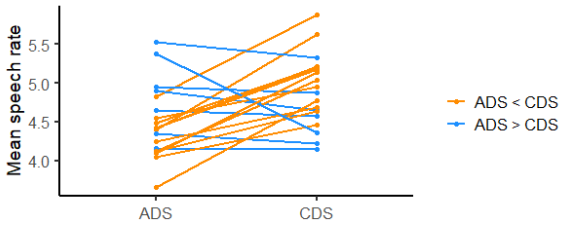

Figure 2: *Individual differences in speech rate as plotted per mother. A similar pattern was observed for articulation rate.*

4. Discussion

This study had three main findings on the speaking rate and fluency in Chinese CDS by examining about 3000 utterances in a semi-spontaneous speech corpus of Chinese ADS and CDS.

First, the most striking finding is that, instead of slowing down, Chinese CDS was actually faster than ADS, particularly for long utterances. The differences in speaking rates between ADS and CDS increased as utterances became longer. The question then arises is: why do Chinese mothers speed up in CDS? On the one hand, CDS utterances are typically shorter than ADS, possibly due to children's limited attention span. Therefore, mothers attempt to condense as much information as possible into each utterance, leading to a faster rate of articulation. On the other hand, CDS often has a positive affect, and happy speech is often characterized by a fast speech rate [36]–[38]. Mothers may increase their speech rate due to their happy emotions, although this might not be the case for other languages like Dutch, where happy speech could be slow [6], [38]. In general, we observed language-specific temporal modifications in Chinese CDS, consistent with previous research using different measures and examining different contexts [5]–[7]. As discussed by Han *et al*. [6], there could be cross-linguistic differences in temporal modifications in CDS across languages with different rhythmic classes.

Studies on the prosodic features of CDS often examine speech samples from mother-child interactions in natural or semi-structured play settings, while ADS samples typically come from conversations or interviews with experimenters. This leads to significant variations in the content and context of the speech data in ADS and CDS. It remains unclear whether the differences in speaking rates observed in these studies were magnified by the contextual and content differences. Here, we used a storybook telling task to elicit semi-spontaneous ADS and CDS, which can maximally match the content and context of speech data between conditions while maintaining the naturalness of speech. Using the same method and materials, we found that Chinese mothers modified their pitch in CDS, and Dutch mothers had a slower speaking rate and a higher pitch in CDS compared to ADS ([6], [39]). Therefore, our results cannot be solely attributed to the speech elicitation method.

If a slow speaking rate is beneficial for language learning (e.g., [24], [25]), then how does a faster speaking rate influence learning? For example, does an increase in articulation rate lead to hypoarticulation in segmental properties? Research has indicated that when speaking at a fast rate (about 6.67 syllables/s), lexical tones in Chinese (ADS) display higher, flatter pitch contours and a smaller tonal space [40]. However, the impact of speaking fast on the lexical tones in Chinese CDS remains unclear because CDS generally has a higher pitch and more pitch variations than ADS [41]. It is possible that a faster speaking rate in CDS and a larger pitch range at the intonational level in CDS could potentially increase lexical pitch range and tonal space. Previous studies on tonal hyperarticulation have not taken speaking rate into account as a factor (e.g., [42]). Future studies should explore tonal hyperarticulation across a wide range of speaking rates and investigate whether the impact of speaking rate on lexical tonal hyperarticulation in CDS varies in tonal languages with different rhythmic classes (e.g., Chinese, a tonal, syllable-timed language, vs. Thai, a tonal, stress-timed language). Moreover, studies have indicated that adults are sensitive to changes in speaking rate during speech [43], yet it remains unclear how children adjust to the dynamic variations of speech in CDS and how such variations can influence children's perception and learning. For example, does a fast speaking rate in Chinese CDS capture more attention? And how does it influence word segmentation and word learning in Chinese children? Future studies should address these questions with more detailed acoustic analyses and assess Chinese children's word learning performance by manipulating the speaking rate in the input.

Second, we found that CDS was more fluent than ADS, which is consistent with previous claims that CDS should be more fluent. Specifically, we observed differences in pausing patterns: CDS had fewer silent pauses and filled pauses, and the pause durations were shorter. Fluent speech may help children's language comprehension. However, it is important to note that there were still many silent and filled pauses in CDS. Further analysis should investigate the positions of these pauses and how they influence children's word learning and prosodic phrasing.

Third, we showed individual differences in the fluency of ADS and CDS. We found significant correlations between pausing behavior but not speaking rate, suggesting that speaking rate in CDS is more adaptive while pausing is more personal. Recent research shows that individuals with higher levels of empathy, greater neuroticism, and lower extraversion tend to speak faster when using child-directed language in broadcasting, compared to when using adult-directed language [15]. Further exploration of related factors, such as mothers' personalities and maternal sensitivity, can provide insight into the individual differences in temporal modifications in children.

5. Conclusions

Contrary to previous findings on Dutch, English, German, and other languages, Chinese mothers do not slow down in CDS when similar content is used to elicit ADS and CDS semi-spontaneous speech; instead, they talk faster when producing long utterances. Overall, Chinese CDS is more fluent in CDS. As a result, we should consider cross-linguistic differences when it comes to slowing down as a common feature of CDS. In conclusion, these findings emphasize the language-specific and individual variations in the temporal aspects of CDS.

6. Acknowledgements

This paper was supported by the Shanghai Planning Project of Philosophy and Social Sciences (2020EYY001), the Fellowship of China Postdoctoral Science Foundation (2020M671041), and the Fundamental Research Funds for the Central Universities (2022ECNU-XWK-XK005).

7. References

[1] M. L. Nencheva and C. Lew-Williams, "Understanding why infant-directed speech supports learning: A dynamic attention perspective," *Dev. Rev.*, vol. 66, p. 101047, Dec. 2022.

[2] M. Soderstrom, "Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants," *Dev. Rev.*, vol. 27, no. 4, pp. 501–532, Dec. 2007.

[3] A. Martin, Y. Igarashi, N. Jincho, and R. Mazuka, "Utterances in infant-directed speech are shorter, not slower," *Cognition*, vol. 156, pp. 52–59, Nov. 2016, doi: 10.1016/j.cognition.2016.07.015.

[4] C. R. Narayan and L. C. McDermott, "Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1272–1281, 2016.

[5] Y. Zhang and Y. Gu, "A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting," *Proc. Annu. Meet. Cogn. Sci. Soc.*, vol. 45, no. 45, 2023, Accessed: Jan. 11, 2024.

[6] M. Han, N. H. de Jong, and R. Kager, "Language specificity of infant-directed speech: Speaking rate and word position in word-learning contexts," *Lang. Learn. Dev.*, vol. 17, no. 3, pp. 221–240, Jul. 2021.

[7] M. Han, N. H. de Jong, and R. Kager, "Infant-directed speech is not always slower: Cross-linguistic evidence from Dutch and Mandarin Chinese," in *Proceedings of the 42nd annual Boston University Conference on Language Development*, Cascadilla Press, 2018, pp. 331–344.

[8] A. Malécot, R. Johnston, and P.-A. Kizziar, "Syllabic rate and utterance length in French," *Phonetica*, vol. 26, no. 4, pp. 235–251, Jul. 1972.

[9] H. R. Bosker, A.-F. Pinget, H. Quené, T. Sanders, and N. H. de Jong, "What makes speech sound fluent? The contributions of pauses, speed and repairs," *Lang. Test.*, vol. 30, no. 2, pp. 159–175, Apr. 2013.

[10] K. Read, E. Furay, and D. Zylstra, "Using strategic pauses during shared reading with preschoolers: Time for prediction is better than time for reflection when learning new words," *First Lang.*, vol. 39, no. 5, pp. 508–526, Oct. 2019.

[11] M. Soderstrom and J. L. Morgan, "Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech," *Dev. Sci.*, vol. 10, no. 5, pp. 641–653, 2007.

[12] C. Kidd, K. S. White, and R. N. Aslin, "Toddlers use speech disfluencies to predict speakers' referential intentions," *Dev. Sci.*, vol. 14, no. 4, pp. 925–934, Jul. 2011.

[13] N. H. de Jong, "Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency," *Int. Rev. Appl. Linguist. Lang. Teach.*, vol. 54, no. 2, pp. 113–132, 2016.

[14] T. M. Derwing, M. J. Munro, R. I. Thomson, and M. J. Rossiter, "The relationship between L1 fluency and L2 fluency development," *Stud. Second Lang. Acquis.*, vol. 31, no. 4, pp. 533–557, Dec. 2009.

[15] Y. Zhang and Y. Gu, "Individual differences in multimodal child-directed language: Unraveling individual style, empathy and the Big Five personality traits," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46, in press.

[16] J. Shi, Y. Gu, and G. Vigliocco, "Prosodic modulations in child-directed language and their impact on word learning," *Dev. Sci.*, vol. 26, pp. 1–16, 2023.

[17] M. Han, N. H. de Jong, and R. Kager, "Relating the prosody of infant-directed speech to children's vocabulary size," *J. Child Lang.*, vol. 51, no. 1, pp. 217–233, 2024.

[18] C. Cox *et al.*, "A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech," *Nat. Hum. Behav.*, vol. 7, no. 1, Art. no. 1, Jan. 2023.

[19] D. Raneri, K. V. Holzen, R. Newman, and N. B. Ratner, "Change in maternal speech rate to preverbal infants over the first two years of life," *J. Child Lang.*, vol. 47, no. 6, pp. 1263–1275, Nov. 2020.

[20] A. Fernald and T. Simon, "Expanded intonation contours in mothers' speech to newborns," *Dev. Psychol.*, vol. 20, no. 1, pp. 104–113, 1984.

[21] J. Sjons, T. Hörberg, R. Östling, and J. Bjerva, "Articulation rate in Swedish child-directed speech increases as a function of the age of the child even when surprisal is controlled for," *ArXiv170603216 Cs*, Jun. 201.

[22] A. R. Bradlow, "Confluent talker-and listener-oriented forces in clear speech production," *Lab. Phonol.*, vol. 7, 2002.

[23] R. M. Uchanski, "Clear speech," *Handb. Speech Percept.*, pp. 207–235, 2005.

[24] J. Y. Song, K. Demuth, and J. Morgan, "Effects of the acoustic properties of infant-directed speech on infant word recognition," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 389–400, Jul. 2010.

[25] R. Zangl and D. L. Mills, "Increased brain activity to infant-directed speech in 6-and 13-month-old infants," *Infancy*, vol. 11, no. 1, pp. 31–62, 2007.

[26] U. Marklund, E. Marklund, F. Lacerda, and I.-C. Schwarz, "Pause and utterance duration in child-directed speech in relation to child vocabulary size," *J. Child Lang.*, vol. 42, no. 5, pp. 1158–1171, Sep. 2015.

[27] P. Lennon, "The lexical element in spoken second language fluency," in *Perspectives on fluency*, University of Michigan, 2000, pp. 25–42.

[28] N. de Jong, "Fluency in second language testing: Insights from different disciplines," *Lang. Assess. Q.*, vol. 15, no. 3, pp. 237–254, Jun. 2018.

[29] E. R. Felker, H. E. Klockmann, and N. H. D. Jong, "How conceptualizing influences fluency in first and second language speech production," *Appl. Psycholinguist.*, vol. 40, no. 1, pp. 111–136, Jan. 2019.

[30] W. Levelt, "Producing spoken language," *Neurocognition Lang.*, pp. 83–122, 1999.

[31] C. Bellinghausen, S. Betz, K. Zahner, A. Sasdrich, M. Schröer, and B. Schröder, "Disfluencies in German adult-and infant-directed speech," in *Proceedings of SEFOS: 1st International Seminar on the Foundations of Speech. Breathing, Pausing and The Voice*, 2019.

[32] M. Han, *The role of prosodic input in word learning*. Netherlands Graduate School of Linguistics, 2019.

[33] P. P. G. Boersma, "Praat, a system for doing phonetics by computer," *Glot Int.*, vol. 5, 2002.

[34] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021.

[35] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, 2015.

[36] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, no. 1, pp. 227–256, Apr. 2003.

[37] G. M. Grimshaw, J. A. Séguin, and H. K. Godfrey, "Once more with feeling: The effects of emotional prosody on hemispheric specialisation for linguistic processing," *J. Neurolinguistics*, vol. 22, no. 4, pp. 313–326, Jul. 2009.

[38] T. Benders, "Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent," *Infant Behav. Dev.*, vol. 36, no. 4, pp. 847–862, Dec. 2013.

[39] M. Han, N. H. de Jong, and R. Kager, "Pitch properties of infant-directed speech specific to word-learning contexts: a cross-linguistic investigation of Mandarin Chinese and Dutch," *J. Child Lang.*, vol. 47, no. 1, pp. 85–111, 2020.

[40] P. Tang and S. Li, "The acoustic realization of Mandarin tones in fast speech.," in *INTERSPEECH*, 2020, pp. 1938–1941.

[41] T. Benders, J. StGeorge, and R. Fletcher, "Infant-directed speech by Dutch fathers: Increased pitch variability within and across utterances," *Lang. Learn. Dev.*, vol. 17, no. 3, pp. 292–325, Jul. 2021.

[42] P. Tang, N. Xu Rattanasone, I. Yuen, and K. Demuth, "Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech," *J. Acoust. Soc. Am.*, vol. 142, no. 5, pp. 2823–2835, Nov. 2017.

[43] H. Quené, "On the just noticeable difference for tempo in speech," *J. Phon.*, vol. 35, no. 3, pp. 353–362, Jul. 2007.