Bridging Word and World: Vocal Iconicity in Chinese Child-Directed Speech and Child Production

Mengru Han (mrhan@zhwx.ecnu.edu.cn), Yiqi Nie (10210110421@stu.ecnu.edu.cn)
Department of Chinese Language and Literature, East China Normal University,
500 Dongchuan Rd. Shanghai, 200241 China

Yan Gu (yan.gu@ucl.ac.uk)
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK,

Abstract

This study examines three types of vocal iconicity—sound effects, onomatopoeia, and iconic prosody—in Chinese child-directed speech (CDS), adult-directed speech (ADS), and child production. We analyzed a corpus of semi-spontaneous ADS and CDS from forty Chinese mother-child dyads, where the children were 18 and 24 months old. Our findings revealed that (1) mothers used significantly more sound effects and iconic prosody, but not onomatopoeias, in CDS compared to ADS. Interestingly, mothers' iconic prosody was also acoustically more congruent with lexical meanings; (2) The frequency of sound effects was lower than iconic prosody but higher than onomatopoeias; and (3) Chinese children aged 18 or 24 months seldom produced onomatopoeia or iconic prosody. These findings suggest that iconicity is more prevalent and prosodically marked in CDS than in ADS, which may help children's word-to-world mapping. Also, iconic prosody is an advanced prosodic skill that is not typically developed by two-year-old children.[1]

Keywords: vocal iconicity; child-directed speech; child production; iconic prosody

Introduction

Iconicity refers to "the perceived resemblance between the form and meaning of a sign." (Radden, 2021). While arbitrariness has traditionally been considered the main characteristic of human language (Saussure, 1989), there is a growing body of evidence indicating that iconicity is more deeply rooted in human communication than previously thought. Iconicity is prevalent in sign languages and can be found in both gestural and vocal modalities in spoken languages (Perniss, Thompson, & Vigliocco, 2010). Research has shown the role of iconic gestures in early language acquisition (Aussems & Kita, 2021; Bohn, Call, & Tomasello, 2016; Namy, Campbell, & Tomasello, 2004), but there is a lack of research on the development of vocal iconicity, particularly in terms of iconic prosody (e.g., looong) (Perlman, Clark, & Johansson Falck, 2015; Perlman, 2024). For example, our understanding of how common vocal iconicity is in early language input and how children develop it is limited. This paper presents a systematic comparison of three types of vocal iconicity between Chinese adult-directed speech (ADS) and child-directed speech (CDS): onomatopoeia, sound effects, and iconic prosody. Additionally, the paper examines the emergence of vocal iconicity in child production.

Vocal Iconicity and Early Word Learning

When children learn new words, they need to map the sounds to their meanings. Most of these mappings appear to be arbitrary, except for onomatopoeias, which are words imitating sounds of animals, objects, or other noises, like "meow" and "buzz" in English. Relatedly, sound effects are the sounds produced when speakers imitate the sounds of an action or an animal, such as the sound of eating or barking. While previous studies often categorize sound effects under the umbrella of onomatopoeias (Motamedi et al., 2021), we further make a clear distinction between the two for the following reasons: First, compared to onomatopoeias, sound effects may be more iconic and easier or funnier to produce (Dingemanse & Thompson, 2020), and second, onomatopoeias are lexicalized whereas sound effects are not, resulting in different lexical representations. Nevertheless, these iconic mappings between form and meaning may bootstrap learning (Imai & Kita, 2014). As Motamedi et al. (2021) proposed, onomatopoeias (including sound effects) may support children's early word learning by allowing them to access the sensory properties of real-world referents. They help children understand the speech sounds they hear, which can refer to objects or actions in the world (Imai & Kita, 2014). In addition, children acquire a lexical repertoire based on onomatopoeia; for example, "choo choo" refers to a train in English, and "miaomiao" refers to a cat in Chinese.

In addition to onomatopoeia and sound effects, recent studies have revealed that iconic prosody is another important element of vocal iconicity. Iconic prosody refers to the use of pitch, speech rate, stress, and rhythm in speech to imitate or reflect the characteristics of what is being described. It is typically shown in semantic dimensions such as spatial position/direction (e.g., up, down, high, low), size (e.g., big, small), amount (all, more, full), speed (quick, fast, slow, lazy),

[1] Preliminary results of this paper have been reported in the *Proceedings of Speech Prosody 2024*. In this paper, we added acoustic analyses of iconic prosody.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

distance (far, close, long, short), loudness (noisy, loud, quiet, asleep), etc. For example, speakers may elongate the word "long" as "looong" or raise the pitch when saying "up." Iconic prosody is rooted in sensorimotor properties (Ćwiek & Fuchs, 2019), and it can aid children in learning words in these dimensions by offering a simulation or embodiment of the meanings (Gu, Li, & Vigliocco, in prep). These words are more abstract than concrete words, which are harder to learn (Gleitman et al., 2005). However, by using iconic prosody to mark these words, their degree of abstractness decreases, making it easier for children to extract them from speech and, therefore, to learn.

In general, children are perceptually biased towards words that have a higher degree of iconicity (Dingemanse et al., 2015; Laing, 2017), and they learn iconic words earlier (Perry et al., 2018). They might employ different forms of vocal iconicity as a foundation for language learning. Nevertheless, the questions remain: how does vocal iconicity manifest in children's language input, and when do children start to use it in their production?

Vocal Iconicity in Language Input

When addressing children, mothers often use a unique speaking style known as child-directed speech (CDS), which plays a crucial role in language acquisition (Soderstrom, 2007). CDS differs from adult-directed speech (ADS) in acoustic features, lexical choices, prosody, syntactic features, etc. (Fernald et al., 1989; Havron et al., 2023; Kuhl et al., 1997). So far, studies that have explored vocal iconicity in CDS have focused on onomatopoeia. Onomatopoeia generally occurs more frequently in CDS than ADS, and the frequency decreases between 18 and 36 months as children's vocabulary size increases rapidly (Fernald & Morikawa, 1993; Motamedi et al., 2021). Many child-directed specific words are in onomatopoeic forms (Ota, Davies-Jenkins, & Skarabela, 2018). Also, they are prosodically more salient compared to conventional words in CDS (Laing, Vihman, & Keren-Portnoy, 2017).

The most prominent feature of CDS is its exaggerated prosody. Recent studies indicate that mothers use prosody to highlight unfamiliar words compared to familiar words (Han, de Jong, & Kager, 2020, 2021; Shi, Gu, & Vigliocco, 2023), and that the prosodic adaptation in CDS can predict children's learning outcomes (Han, de Jong, & Kager, 2024; Shi, Gu, & Vigliocco, 2023). Furthermore, mothers adjust their use of CDS based on the age of their children (e.g., (Kitamura & Burnham, 2003). Using a shared-book reading task, Herold, Nygaard, and Namy (2012) measured the prosody of dimensional adjectives (e.g., big, small, hot, cold) in CDS and found that mothers modulate amplitude and duration to distinguish dimensional adjectives. This study used a specific set of target words to elicit contrastive adjectives in CDS rather than examining how mothers naturally incorporate iconic prosody in CDS.

Vocal Iconicity in Child Production

Cross-linguistically, onomatopoeias are among the first words that young children produce (Laing, 2014; Tardif et al., 2008). Perry et al. (2018) examined the relationship between frequency and iconicity for about 2000 English words and discovered that younger children tend to use more iconic words. In relation to iconic vocal production, children start to produce iconic gestures relatively late, usually around 26 months of age (Özçalışkan & Goldin-Meadow, 2011), although children who speak verb-biased languages such as Turkish master them as early as 19 months (Furman, Küntay, & Özyürek, 2014). It seems that despite children's perceptual bias towards learning iconic forms, producing them is a more advanced skill that is acquired later in life. As no study has specifically focused on the onset of iconic prosody in children before 24 months of age, though see Gu et al. (in prep) for an ongoing study of English-learning children aged 24 to 52 months, its developmental trajectory remains unclear.

Current Study

To sum up, examining vocal iconicity in CDS will reveal the iconic language environment a child is exposed to. Also, understanding the emergence of iconicity in language production will provide us with a window into children's early language and cognitive development. Importantly, previous studies on the role of iconicity in child language acquisition have predominantly been done on English-learning children. However, languages (especially non-Indo-European languages) differ vastly in the prevalence of vocal iconicity (e.g., sound symbolism) (Imai & Kita, 2014). The aim of this study is to better understand the use of three forms of vocal iconicity—sound effects, onomatopoeia, and iconic prosody—in Chinese CDS and child production. Specifically, we ask the following research questions:

(1) What are the differences in the frequency of vocal iconicity between CDS and ADS, and what is the distribution of the three types of iconicity in CDS? Based on previous studies indicating that caregivers tend to use more iconic gestures (Campisi & Özyürek, 2013; Zhang & Gu, 2023) or iconic signs (Perniss et al., 2018) in child-directed language than adult-directed language, we predict that mothers use more vocal iconicity in CDS than in ADS. Also, previous research has shown that onomatopoeia is frequent in both child production and CDS (see Laing et al., 2017 for a review) and iconic prosody is rare (Motamedi et al., 2024). Therefore, we hypothesize that mothers will use more onomatopoeias (and sound effects) than iconic prosody.

(2) Does the frequency of vocal iconicity change in CDS addressed to 18- and 24-month-old children? Based on previous research on the age-related changes of CDS, it has been shown that the use of onomatopoeias by English mothers decreases, while their use of iconic gestures increases as children get older (Motamedi et al., 2021; Namy et al., 2004). Thus, we predict that Chinese mothers' use of onomatopoeia and sound effects will decrease between 18 and 24 months. The trajectory for iconic prosody is less clear, as it is more abstract. If it follows a pattern similar to that of

iconic gestures, it will increase with children's age. However, a recent study did not show an age effect (Motamedi et al., 2024).

(3) How is iconicity prosodically modulated in CDS? Here, we combined subjective judgments of iconic prosody with acoustic analyses to compare the acoustics of items with iconic prosody compared to those same items that are not judged as iconic. Our prediction is that items perceived as using iconic prosody will be prosodically modulated to convey the intended meaning, and this should also be reflected in the objective acoustic realizations.

(4) Given the lack of research on children's vocal iconicity production, we ask if 18- and 24-month-old children begin to produce the three types of vocal iconicity, especially iconic prosody. Since onomatopoeia is the most common word class in children's early lexicon and iconic gestures emerge later, we predict that children begin producing onomatopoeias and sound effects at 18 and 24 months before producing iconic prosody.

Method

Speech Corpus and Participants

To address the research questions, we analyzed a corpus of ADS and CDS (Han, 2019). Participants included forty Mandarin-Chinese-speaking mother-child dyads. The child participants were divided into two age groups: 18-month-old (N = 21, 9 girls and 12 boys; mean age = 18;15, age range = 17;21–18;27) and 24-month-old (N = 19, 10 girls and 9 boys, age range = 23;27–24;30). All children were typically developing and had no reported language impairments or hearing problems.

A storybook was created to elicit ADS and CDS. It consisted of 12 pages, with each page featuring a word on the left side and an illustration of the word on the right side. Mothers were free to construct the story, as no other script was provided besides the words. Mothers were required to include the words given on each page. As such, both ADS and CDS were semi-spontaneously speech.

Each participant mother told 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, the mothers told the same story to an adult (female, a Mandarin native speaker), while the child was not present. The order of the two speech registers was counterbalanced across participants. Since the mothers told the same story in both ADS and CDS, we can compare their adaptation when the content was similar in the two speech registers.

Data Coding

Speech Transcription and Annotation We used an automatic Chinese speech recognition tool developed by Iflytek (https://www.iflyrec.com) to transcribe speech data in the corpus, which was then divided into utterances by two Chinese speakers, resulting in a total of 6740 utterances. Following Martin et al. (2016), utterances were 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)". In ADS, the speech of the mother was transcribed, while in CDS, the speech of both mothers and children was transcribed. All utterances were manually aligned with the speech data in Praat, and a third native speaker listened through each utterance to verify the accuracy of the transcription.

Onomatopoeia and Sound Effect Each utterance was coded for the presence of onomatopoeia (1 = with onomatopoeia; 0 = no onomatopoeia) and sound effects (1 = with sound effect; 0 = no sound effect) in Praat. A second coder went through all the data, and the intercoder reliability was 100%.

Iconic Prosody Iconic prosody was subjectively judged by three Chinese L1 speakers, as was done in Motamedi et al. (2024). We created a list of Chinese words that could elicit iconic prosody, including seven abstract dimensions: amount (e.g., *duo* 'many'), directions/position (e.g., *gao* 'high'), distance (e.g., *yuan* 'far'), loudness (e.g., *qing* 'quiet' and whispering), size (e.g., *da* 'big'), speed (e.g., *kuai* 'quick'), and strength (*shijin* 'hard') (Motamedi et al., 2024). Additionally, we included verbs that denote upward and downward movement, as well as fast or slow movement (e.g., *tiao* 'jump') (Ekström, Nirme, & Gärdenfors, 2022). Next, a trained research assistant (a native speaker) listened to all the utterances containing target words and marked a "1" whenever there was iconic prosody produced by the mother or the child in Praat (Boersma & Weenink, 2022). The first author listened to all the target words and identified those with agreements. Subsequently, the first author and the third author listened through all the utterances marked with iconic prosody and included only those on which they both agreed (1 = with iconic prosody; 0 = no iconic prosody). Thus, only the items that were agreed upon by three native speakers were included in the analysis. Intercoder agreement reached 93.33%, with a Cohen's Kappa of 0.87, indicating strong consistency. For these items, we further coded their dimensions.

In addition to subjective judgments, we also conducted acoustic analyses of iconic prosodic modulation. As there is a debate on the interaction between lexical tones in Chinese (pitch at the lexical level) and iconic pitch (Ying, 2009), our acoustic analysis focused on the temporal aspect of iconic prosody, namely speaking rate. Specifically, we tested whether there was a difference in speaking rate between items that were judged as iconic and those that were not (e.g., comparing the speaking rate of "long" versus "looong"). We identified eighteen different word types of abstract dimensions that displayed at least one instance of iconic prosody. Thirteen word types displayed iconic prosody in relation to speaking rate. Then we coded their speech rate in correspondence with the original word meanings (fast/slow). For example, the word *kuai* ('quick') was coded as being associated with a "fast" speaking rate, and *chang* ('long') was coded as a being linked with a "slow" speaking rate. Then we

manually marked all occurrences of these word types (N = 560) in the corpus and measured their speech rate (syllables/s) using a Praat script. Following the subjective coding mentioned above, they were then further categorized into iconic (N = 63) and non-iconic items (N = 497).

Data Analysis

We used R (R Core Team, 2021) for data analysis. First, we compared the frequency of vocal iconicity across speech registers (ADS/CDS) and ages (18 months/24 months). The response variable was a binary code indicating whether an utterance contained each type of vocal iconicity (if yes, coded as 1; otherwise, 0). To compare the differences in the proportion of vocal iconicity, we used a generalized linear mixed model (GLMM) implemented through the 'glmer' function from the 'lme4' package (Bates et al., 2015). The fixed effects were Speech Register (ADS/CDS), Age (18 m/24 m), as well as their interaction. The interaction between Speech Register and Age was dropped when it did not significantly improve the model. "Participant" was added as a random intercept.

Second, to compare the number of dimensions between ADS and CDS, we used the Wilcoxon signed-rank test.

Third, we used a linear mixed-effects model implemented through the 'lmer' function from the 'lme4' package for the acoustic analysis of iconic prosodic modulation (Bates et al., 2015). The dependent variables were speech rate (syllables/s), which was log-transformed before data analysis. The fixed effects were Iconicity (iconic/non-iconic) and Prosodic Modulation (e.g., fast/slow), as well as their interaction. Speech Register (ADS/CDS) was also controlled for in the model. The random effects were Participant and Word type.

Results

Vocal Iconicity in ADS and CDS

The Proportions of Vocal Iconicity between ADS and CDS We first compared the proportions of the three types of vocal iconicity between ADS and CDS, as well as between the two age groups. We did not observe any age-related changes in mothers' use of vocal iconicity, as the effect of age was not significant for any of the three types. As such, the effect of age was removed from the models. When comparing the two speech registers (Figure 1), the results showed that mothers used significantly fewer sound effects (*beta* = -1.05, p = 0.019) and iconic prosody (*beta* = -0.91, p = 0.003) in ADS (N_{SE} = 6; N_{IP} = 13) compared to CDS (N_{SE} = 36; N_{IP} = 75). However, there was no difference in the proportions of onomatopoeia between the two speech registers (p = 0.31, N_{ADS} = 4; N_{CDS} = 18).

Furthermore, iconic prosody was significantly more frequent than sound effects (W = 78.5, p = 0.001), which in turn were more frequent than onomatopoeias (W = 32.0, p = 0.011). These findings indicate that iconic prosody, although understudied in previous research, seems to be the most preferred form of vocal iconicity, at least in this Chinese CDS corpus.

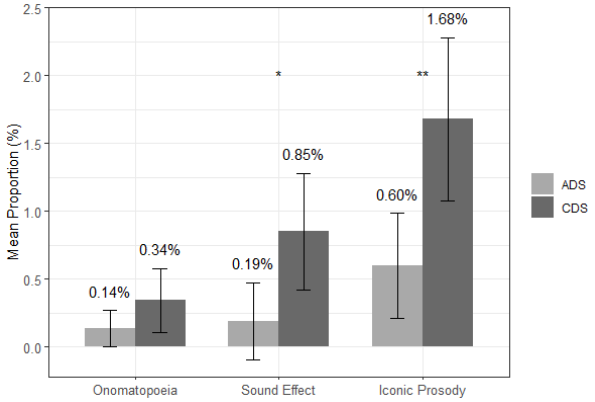

Figure 1: Mean proportions (95% CI) of onomatopoeia, sound effects, and iconic prosody in ADS and CDS.

Iconic Prosodic Modulation Figure 2 shows boxplots of speech rate categorized by subjectively judged iconicity (iconic/non-iconic) and speech modulation (fast/slow). After controlling for speech register (*beta* = 0.02, p = 0.49), the acoustic analysis of speech rate modulation revealed that both iconicity (*beta* = -0.29, p = 0.007) and speech rate modulation (*beta* = -0.69, p < 0.001) were significant. Additionally, there was a significant interaction between iconicity and speech rate modulation (*beta* = 0.72, p < 0.001, Figure 2), indicating that the speech rate was only slowed down when the word was subjectively judged as iconic (*beta* = -0.69, p < 0.001), but not when judged as non-iconic (*beta* = 0.04, p = 0.82). As such, the words judged with iconic prosody showed a significant difference in speaking rate between fast and slow concepts, whereas those judged to be non-iconic did not show such a contrast. This supports our subjective judgment of iconic prosody and indicates that mothers indeed adjust their speech rate to mark iconicity regardless of speech register.

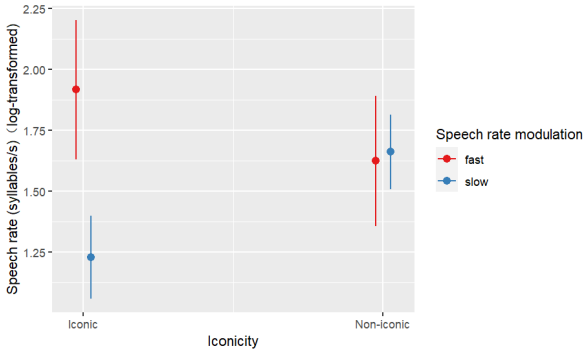

Figure 2: Predicted values of speech rate (syllables/s, log-transformed) as a function of subjectively-judged iconicity (Iconic/Non-iconic) and word concepts (Fast/Slow).

The Dimensions of Iconic Prosody in CDS As shown in Figures 3, in general, CDS had a broader range of dimensions compared to ADS: iconic prosody in ADS was restricted to amount, direction, distance, motion, and size, while CDS included seven dimensions.

When aggregating the number of dimension types per condition per participant, the Wilcoxon signed-rank tests revealed that both at 18 and 24 months, CDS (18m: $M = 1.14$, $SD = 1.06$; 24m: $M = 1.26$, $SD = 1.59$) had significantly more dimensions than ADS (18m: $M = 0.19$, $SD = 0.4$; $p = 0.002$; 24m: $M = 0.37$, $SD = 0.6$, $p = 0.022$). This suggests that there are more types of dimensions in CDS compared to ADS. Additionally, a Mann-Whitney U test showed that there was no significant effect of age on the number of dimensions ($p = 0.8$), suggesting that there was no evidence of age-related changes in the number of dimension types in this six-month gap.

In summary, besides the more frequent occurrence of iconic prosody in CDS, this speech register also has a wider variety and a greater number of dimensions of iconic prosody.

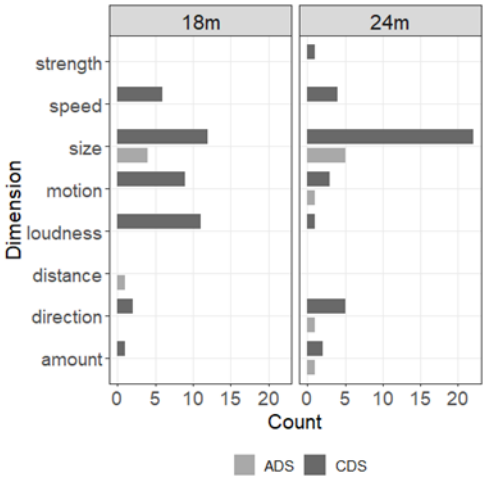

Figure 3: Number of cases with iconic prosody (as divided by dimension) in ADS and CDS.

Vocal Iconicity in Child Production

During the CDS condition, children had a total of 531 speech productions (including utterances and verbal productions that were unclear). There was no significant effect of age on the proportions of the three types of vocal iconicity. As shown in Figure 4, the most common type of child production was sound effects, with a total of 13 cases, accounting for 2.45% of all child productions. However, none of these productions included onomatopoeia, the lexicalized form of natural sounds. Interestingly, there was a single instance of iconic prosody at 24 months: using an extremely high pitch when producing "high". It should be noted that in this case, the child repeated the mother's iconic high-pitched prosody. This suggests that children may not start using iconic prosody until at least 24 months of age.

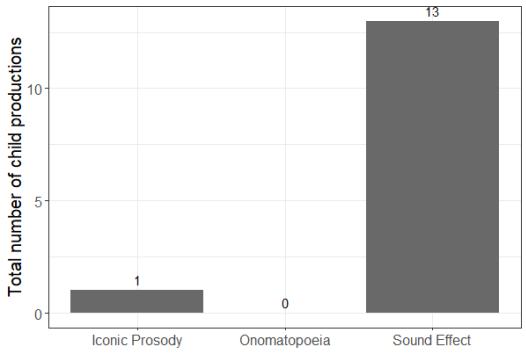

Figure 4: Onomatopoeia, sound effects, and iconic prosody in 18- and 24-month-old children's production.

Discussion

The current study examines the development of vocal iconicity in language acquisition. Specifically, we conducted a systematic comparison of the use of vocal iconicity in CDS and ADS with similar contents and investigated how vocal iconicity emerges in young children. Importantly, we extended previous research to include iconic prosody, a type of vocal iconicity that has received little attention in existing literature.

As predicted, CDS is generally more iconic than ADS. In particular, mothers employed a higher frequency of sound effects and more instances of iconic prosody in CDS as compared to ADS. Furthermore, CDS had a greater variety and more dimensions of iconic prosody than ADS. These findings suggest that CDS is tailored to the developmental stage of children, with mothers employing more iconic features that could potentially enhance language learning.

By combining perception and acoustic analysis, we have demonstrated that mothers naturally incorporate iconic prosodic modulation in speech. While previous studies indicate that Chinese CDS does not show any evidence of slowing down compared to ADS (Han et al., 2021; Zhang & Gu, 2023), our findings suggest that they do manipulate temporal cues to highlight iconicity. They adjust their speech rate to convey corresponding word meanings in relation to fast and slow, potentially enhancing children's perception and cognition of concepts related to fast and slow. This provides embodied input for children to map sounds to their meanings.

The current acoustic analysis of iconic prosody focuses on temporal measures. While most languages use a high pitch to convey smallness (e.g., Ohala, 1983), in Chinese, words associated with smallness, such as *xiao* ("small") and *ai* ("short"), tend to have a low-dipping tone. It is possible that the iconic prosodic modulation of smallness in Chinese may be a low pitch. There is still debate about how iconic prosody is used to convey smallness and largeness in Chinese and how it interacts with lexical tonal pitch. Future analysis should consider using a more detailed measure to account for factors such as lexical tones.

While onomatopoeias are often considered to be words that children learn early on and are more prevalent in CDS compared to ADS, our findings indicate that the frequency of lexicalized onomatopoeias is not significantly higher in CDS than in ADS. Instead, it is the sound effects that are significantly more frequent in CDS. It remains to be seen whether a similar pattern exists in other languages. Furthermore, despite previous research suggesting that onomatopoeias are among the first words uttered by children, the children in this specific study, aged 18 and 24 months, did not produce any onomatopoeic words. Instead, they produced sound effects during mother-child interaction. It is important to note that previous research did not differentiate between sound effects and onomatopoeia, so it would be beneficial to explore these distinctions in other studies and other languages. Also, since studies on early vocabulary often rely on parental reports, such as the M-CDI, it is possible that parents may not differentiate between these two types of lexical representations. This suggests that future research should differentiate between onomatopoeia and sound effects when studying early lexical development, as they may have different roles in language production. In addition, we used subjectively judged iconic prosody in the current study. Further acoustic analyses are required to determine the salience of items with iconic prosody compared to those same items that are not judged as iconic.

While onomatopoeias are often considered to be words that children learn early on and are more commonly used in CDS compared to ADS, our findings indicate that the frequency of lexicalized onomatopoeia is not significantly different between the two speech registers. Instead, it is the sound effects that are significantly more frequent in CDS. Additionally, despite previous research considering onomatopoeia as the first words uttered by children, in this particular corpus, children at 18 and 24 months of age did not produce any lexicalized onomatopoeia, but they did produce sound effects in this task. Since studies on early vocabulary often rely on parental reports such as the M-CDI (Fenson et al., 2007), it is possible that parents may not distinguish between these two different types of lexical representations when answering questions. This suggests that future research should differentiate between onomatopoeia and sound effects when studying early lexical development.

Conclusions

In conclusion, iconicity is more prevalent and variable in CDS than in ADS, and iconic prosody is an advanced prosodic skill that is typically not developed by two-year-old children. Moreover, children are able to produce sound effects, but they do not yet produce lexicalized onomatopoeias by this age. Despite language being predominantly arbitrary, speakers, especially caregivers, use iconicity to bridge the mapping between words and the world. Consequently, the significance of vocal iconicity, particularly iconic prosody, in language acquisition calls for further investigation.

Acknowledgements

We thank the reviewers of CogSci 2024 for their comments and suggestions. 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).

References

Aussems, S., & Kita, S. (2021). Seeing iconic gesture promotes first- and second-order verb generalization in preschoolers. *Child Development*, *92*(1), 124–141. https://doi.org/10.1111/cdev.13392

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (6.2.05) [Computer software]. http://www.praat.org/

Bohn, M., Call, J., & Tomasello, M. (2016). Comprehension of iconic gestures by chimpanzees and human children. *Journal of Experimental Child Psychology*, *142*, 1–17.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14–27.

Ćwiek, A., & Fuchs, S. (2019). Iconic prosody is rooted in sensori-motor properties: Fundamental frequency and the vertical space. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 1572–1578.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Dingemanse, M., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. *Language and Cognition*, *12*(1), 203–224. https://doi.org/10.1017/langcog.2019.49

Ekström, A. G., Nirme, J., & Gärdenfors, P. (2022). Motion iconicity in prosody. *Frontiers in Communication*, *7*.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). *The MacArthur Communicative Development Inventories: User's guide and technical manual*.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & 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

Furman, R., Küntay, A. C., & Özyürek, A. (2014). Early language-specificity of children's event encoding in speech and gesture: Evidence from caused motion in Turkish.

Language, Cognition and Neuroscience, *29*(5), 620–634. https://doi.org/10.1080/01690965.2013.824993

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Gu, Y., Li, M., & Vigliocco, G. (In prep). The development of iconic prosody in children aged 2-4 years old.

Han, M. (2019). *The role of prosodic input in word learning*. Netherlands Graduate School of Linguistics.

Han, M., de Jong, N. H., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language*, *47*(1), 85–111. https://doi.org/10.1017/S0305000919000813

Han, M., de Jong, N. H., & Kager, R. (2021). Language specificity of infant-directed speech: Speaking rate and word position in word-learning contexts. *Language Learning and Development*, *17*(3), 221–240. https://doi.org/10.1080/15475441.2020.1855182

Han, M., de Jong, N. H., & Kager, R. (2024). Relating the prosody of infant-directed speech to children's vocabulary size. *Journal of Child Language*, *51*(1), 217–233. https://doi.org/10.1017/S0305000923000041

Havron, N., De Carvalho, A., Babineau, M., Barbir, M., Dautriche, I., & Christophe, A. (2023). There might be more to syntactic bootstrapping than being pragmatic: A look at grammatical person and prosody in naturalistic child-directed speech. *Journal of Child Language*, 1–5. https://doi.org/10.1017/S0305000923000181

Herold, D. S., Nygaard, L. C., & Namy, L. L. (2012). Say it like you mean it: Mothers' use of prosody to convey word meaning. *Language and Speech*, *55*(3), 423–436.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, *34*(5), 387–405.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, *161*, 32–45. https://doi.org/10.1016/j.jecp.2017.03.017

Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language*, *44*(5), 1117–1139.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, *156*, 52–59. https://doi.org/10.1016/j.cognition.2016.07.015

Motamedi, Y., Murgiano, M., Grzyb, B., Gu, Y., Kewenig, V., Brieke, R., Donnellan, E., Marshall, C., Wonnacott, E., Perniss, P., & Vigliocco, G. (2024). Language development beyond the here-and-now: Iconicity and displacement in child-directed communication. *Child Development*. https://doi.org/10.1111/cdev.14099

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science*, *24*(3), e13066.

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57.

Ohala, J. J. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*, *40*(1), 1–18. https://doi.org/10.1159/000261678

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science*. https://doi.org/10.1111/cogs.12628

Özçalışkan, Ş., & Goldin-Meadow, S. (2011). Is there an iconic gesture spurt at 26 months. *Integrating Gestures: The Interdisciplinary Nature of Gesture. Amsterdam, NL: John Benjamins*.

Perlman, M. (2024). Iconic prosody is deeply connected to iconic gesture, and it may occur just as frequently. In *Handbook on Iconicity in Language*. Oxford: Oxford University Press.

Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic prosody in story reading. *Cognitive Science*, *39*(6), 1348–1368. https://doi.org/10.1111/cogs.12190

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2018). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science*, *21*(2), e12551. https://doi.org/10.1111/desc.12551

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), e12572.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radden, G. (2021). Iconicity. In *The Routledge Handbook of Cognitive Linguistics* (pp. 268–296). Routledge.

Saussure, F. de. (1989). *Cours de linguistique générale*. Otto Harrassowitz Verlag.

Shi, J., Gu, Y., & Vigliocco, G. (2023). Prosodic modulations in child-directed language and their impact on word learning. *Developmental Science*, *26*, 1–16. https://doi.org/10.1111/desc.13357

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Ying, X. (2009). An investigation of the phonetic and semantic resemblance of monosyllabic antonyms in modern Chinese. *Language Teaching and Research*, *3*, 21–27.

Zhang, Y., & Gu, Y. (2023). A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). https://escholarship.org/uc/item/17k7h7m6