# Rhythm and Expression in *The Cat in the Hat*

*Mara Breen*[1], *Sarah Weidman*[1], *Katharine Guarino*[2]

[1] Department of Psychology and Education, Mount Holyoke College, South Hadley, MA, USA
[2] Department of Psychology, The University of Texas at Austin, Austin, TX, USA
mbreen@mtholyoke.edu, weidm22s@mtholyoke.edu, katieguarino22@gmail.com

## Abstract

In recent years, there has been increasing interest in whether rhythmic interventions support young children's literacy development [1]. To begin to explore this connection, we assessed several aspects of rhythmicity and expressivity of productions of the notably rhythmic and rhyming children's book, *The Cat in The Hat* by Dr. Seuss. Participants subjectively rated either the rhythmicity or expressivity of speech taken from recordings of the book read aloud. These perceptual ratings were correlated with acoustic measures of rhythmicity and expressivity. Moreover, we observed a surprising lack of consistency between perceptual ratings of rhythmicity and expressivity. However, we observed a consistent relationship between the perceptual ratings of the first couplet of verses and the second, suggesting that readers showed self-entrainment in rhythmicity and expressivity between verse couplets. These findings can inform our investigation of the role of rhythm in literacy development and set the stage for further investigation into rhythmic entrainment across speakers.

**Index Terms**: rhythm, expression, production, child-directed speech

## 1. Introduction

### 1.1. Rhythm in speech perception

Speech researchers have long been interested in what factors contribute to isochronous speech in which stresses are perceived at regular intervals [2]. Although consistent acoustic measures of isochrony have been difficult to identify [3], there is consensus that perceptual isochrony is an important component of speech perception. For example, recent studies demonstrate that perceptual isochrony can help resolve ambiguity in speech perception [4,5]. Moreover, it is clear that babies attend to rhythm from a very early age [6], and can use rhythmic cues to learn words [7].

Many have observed that children's nursery rhymes are rhythmic and have consistent rhyming schemes. Moreover, exposure to rhyming texts has been shown to improve children's phonological awareness [8,9] and expressive vocabulary abilities [10]. However, little prior work has examined whether and how the rhythmic schemes of children's books improves children's literacy development.

There is reason to suspect that the rhythmicity of these texts supports reading acquisition, though this topic has received little direct investigation. For example, dyslexic children have been shown to be less sensitive to rhythm than normally reading children [11,12]. In addition, reading researchers have long advocated that rhythmic reading, including the use of an Interactive Metronome program, is an effective instruction tool [13]. One possibility, that we begin to explore here, is that rhythmic texts support self-entrainment [14], such that, within a verse, readers produce stresses at a consistent rate.

What can certainly not be denied is the fact that children like to encounter rhythm when reading, as evidenced by the popularity of books by Dr. Seuss and others. The goal of the current study is to explore the nature of the structure of rhythmic texts in order to begin to understand what aspects of their composition might support reading acquisition.

### 1.2. Rhythm in child-directed speech

A parallel line of inquiry in speech comprehension is determining the circumstances under which speech is most likely to be rhythmic. Recent investigations have demonstrated that multiple factors can contribute to the rhythmicity of speech, including the syntactic/semantic structure of the speech [15,16], the cognitive load of the speaker [17], and the audience to which the speech is directed.

It is this final factor that serves as the starting point for the current study. Recent work has demonstrated that the same speakers produce more rhythmic (even-timed) child-directed speech and singing than adult-directed speech [18,19] suggesting that readings of *The Cat in the Hat* will be good candidates for investigating speech rhythmicity.

### 1.3. Expression in child-directed speech

It is clear that expression plays a large role in child-directed speech. Acoustically, the greater perceived expressivity of child-directed speech is realized primarily by higher, and more variable, F0 than adult-directed speech [20]. Moreover, there is evidence that this expressivity facilitates speech comprehension. Examinations of child-directed speech indicate that exaggerated pitch peaks at the ends of utterances marking focused words may facilitate speech processing [21]. Moreover, previous research has demonstrated that infants are able to distinguish words from syllable sequences spanning word boundaries after exposure to infant-directed speech but not after exposure to adult-directed speech [22].

The current study further investigates the role of expression in children's literacy development, specifically examining how expressivity of child-directed speech may relate to the rhythmicity of child-directed speech.

### 1.4. Current Study

The objective of the current study was to investigate the rhythmic and expressive characteristics of child-directed speech; specifically, the book *The Cat in the Hat* by Dr. Seuss. The research questions we addressed were:



10.21437/SpeechProsody.2014-81

(1) Is there a consistent relationship between perceived rhythmicity and perceived expressivity in productions of *The Cat in the Hat?*

(2) Does the rhythmicity of the first couplet of a verse predict the rhythmicity of the second couplet?

(3) Does the expressivity of the first couplet of the verse predict the expressivity of the second couplet?

We could imagine two possible answers to question (1). On the one hand, if greater expressivity and greater rhythmicity are both predictive of better outcomes for audiences (of children), then it could be the case that these two factors will be correlated such that greater perceived rhythmicity corresponds to greater perceived expressivity. On the other hand, rhythmic variation is one marker of expressiveness in production. Therefore, we may observe a negative correlation between perceived rhythmicity and perceived expressivity of verses such that greater perceived rhythmicity corresponds to lower perceived expressivity.

Regarding (2) & (3), we were interested in whether there is rhythmic and expressive consistency within verses after accounting for within-speaker factors. That is, are there characteristics of the verses themselves that predict how expressive or rhythmic their productions will be?

# 2. Method

## 2.1. Participants

Participants were sixteen native speakers of American English (all female), ages 18-35. All received course credit for participating.

## 2.2. Materials

Participants all read "The Cat in the Hat" by Theodor Seuss Geisel (aka "Dr. Seuss") in its entirety. This text was written in an anapestic tetrameter such that every couplet (1A, 1B) consists of four anapestic (weak-weak-**strong**) feet. Two couplets form a verse (1). There are a total of 71 verses in the book.

(1) *A: "Put me **down**!" said the **fish**.*
*This is **no** fun at **all**!*
*B: Put me **down**!" said the **fish**.*
*"I do **not** wish to **fall**."*

## 2.3. Procedure

Participants were randomly assigned to read the hardcover book aloud in one of two environments: In a quiet room in the lab with no audience present; or to an audience of between three and five 5- and 6-yr-olds in a pre-kindergarten classroom. Both groups were instructed to read as naturally as possible. Their productions were recorded with Praat [23] through a Shure SM10A head-mounted microphone, connected to a Rolls Mini-Mic pre-amplifier. Productions were recorded with a sampling rate of 44100 Hz.

Initial analysis of acoustic measures demonstrated no significant differences between the productions recorded in the lab and those recorded in the classroom. This may be because our participants were all familiar with *The Cat in the Hat* and couldn't help but read it expressively and rhythmically regardless of whether there was an audience present. We will, therefore, report data from all productions together.

## 2.4. Acoustic Measurements

All productions were aligned with the book text in Praat using the Prosodylab-Aligner software [24] and then hand-corrected. Productions were segmented into verses (as in 1) and further divided into (A) and (B) couplets. All verses which contained a disfluency in either of the two couplets were excluded from further analysis. Of the 71 total verses in the book, we selected 18 for further analysis which met the following criteria: First, they were primarily anapestic tetrameter (though the first foot could be iambic); second, they appeared on four adjacent lines on one page of the book; third, they were fluently produced by at least 14 of the 16 speakers. These constraints resulted in a total of 252 verse productions (504 couplets).

These couplets were distributed across three lists which were balanced, as much as possible, across speakers and verses. The couplets that comprised each verse (e.g., (1A) & (1B)) did not appear on the same list, so that each couplet was rated by different listeners. The lists were comprised of 173, 166, and 165 couplets, respectively.

Using Praat, we extracted measures of duration, average F0, maximum F0, minimum F0, and average intensity from every word. For the current study, we looked only at measures of duration and mean F0.

## 2.5. Perceptual Evaluation

An additional 30 female participants were recruited to participate in a perceptual evaluation task. None had participated in the production experiment. They were randomly assigned to rate either the expressivity or the rhythmicity of the verse couplets.

Half of the participants (n= 15) rated their perception of the each couplet's rhythmicity on a scale of 1-4 where 1 was "very rhythmic" and 4 was "not at all rhythmic." Instructions were adapted from [5] such that raters were asked to determine how well they could find a steady beat for the couplet. Each of these 15 participants was assigned to one of the three lists, so that five participants rated each couplet. They were first presented with 8 practice couplets; four were categorized by the authors as "very rhythmic"; four were categorized as "not at all rhythmic." We observed agreement on these ratings which was significantly greater than chance, Fleiss' $\kappa = .09$, $z = 10.5$, $p = 0$ [25], indicating that raters were sensitive to couplet rhythm. Ratings from all five raters were averaged, resulting in an average Rhythmicity score for each couplet.

Half of the participants (n=15) rated their perception of the each couplet's expressivity on a scale of 1-4 where 1 was "very expressive" and 4 was "not at all expressive." Each of these 15 participants was assigned to one of the three lists, so that five participants rated each couplet. They were first presented with 8 practice couplets; four were categorized by the authors as "very expressive"; four were categorized as "not at all expressive." We observed agreement on these ratings which was significantly greater than chance, Fleiss' $\kappa = .29$, $z = 34.5$, $p = 0$, indicating that raters were sensitive to couplet expressivity. Ratings from all five raters were averaged, resulting in an Expressivity score for each couplet.

# 3. Results

## 3.1. Acoustic and perceptual ratings

Before investigating our research questions, we measured the consistency between acoustic and perceptual ratings of expressivity and rhythmicity.

### *3.1.1. Acoustic and perceptual measures of rhythmicity*

In order to objectively quantify the rhythmicity of productions, we identified the timing of the onsets of strong syllables (**bolded** in (1)): For each couplet, we computed the standard deviation of the latencies (SD_Latency_Strong) between the onset of strong syllables 1 and 2 (***down*** and ***fish***), syllables 2 and 3 (***fish*** and ***no***), and syllables 3 and 4 (***no*** and ***all***). SD Latency ranged from 1 ms to 850 ms. This measure was highly correlated with the average Rhythmicity rating, $R$ = .48, $p$ < .0001, such that higher variability in the timing onsets of strong syllables was associated with lower Rhythmicity ratings (Figure 1). Recall that a lower number indicates a higher rating of Expressivity. This result suggests that timing variability was one factor that raters used in their Rhythmicity judgments.

### *3.1.2. Acoustic and perceptual measures of expressivity*

As intonational variability has been identified as a strong marker of expressivity in speech [20], and child-directed speech exhibits greater pitch variability [21], we computed an acoustic measure of expressivity based on pitch. For each couplet (e.g., 1A, 1B), we computed the standard deviation of the mean F0 (SD_F0_Word) for each word. This measure was highly inversely correlated with the average Expressivity rating, $R$ = -0.18, $p$ < .001 (Figure 2). Recall that a lower number indicates a higher rating of Expressivity. This result demonstrates that F0 variability was one factor that raters used in their Expressivity judgments.

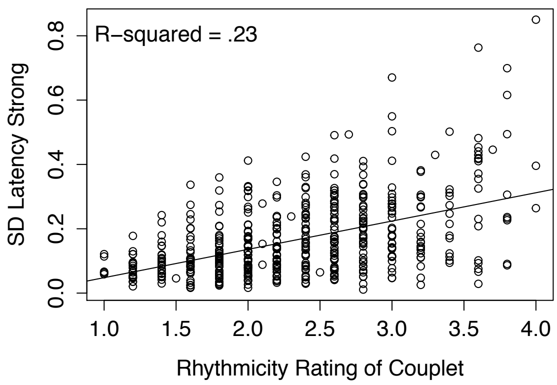


Figure 1: *Correlation of average perceived rhythmicity rating of a couplet and the standard deviation of the latency of onsets between the strong syllables of the couplet.*

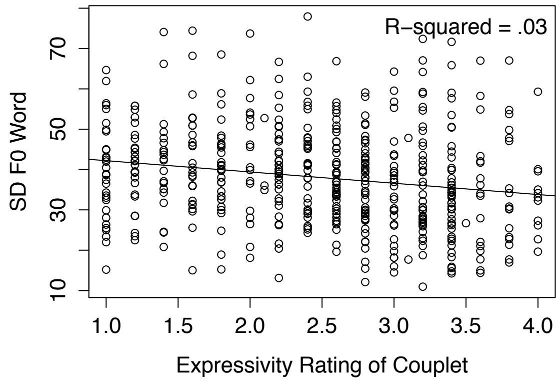


Figure 2: *Correlation of average perceived expressivity rating of a couplet and the standard deviation of the mean F0 of each word in the couplet.*

## 3.2. Relationship between Expressivity and Rhythmicity

In order to determine whether there is a consistent relation ship between rhythmicity ratings and expressivity ratings, we conducted a correlational analysis comparing the Expressivity and Rhythmicity scores for each couplet. Pearson's correlational coefficient was $R$ = .03, which did not reach significance, $p$ = .48, indicating no systematic relationship between perceived expressivity and perceived rhythmicity.

## 3.3. Intraverse Expressivity and Rhythmicity

In order to investigate the consistency between the Rhythmicity and Expressivity scores of the first and second couplets of a single verse, we analyzed these scores separately on a single-couplet basis using a linear mixed-effects model implemented in the languageR package [26] implemented in the R statistical programming language (R Core Development Team, 2012).

We computed models in which we predicted an A couplet's Rhythmicity/Expressivity score from one fixed effect (the B couplet's Rhythmicity/Expressivity score) and two random effects: The individual reader and the verse that the couplets comprised. In this way, we could see whether the Rhythmicity/Expressivity of one half of a couplet predicted the Rhythmicity/Expressivity of the second half after accounting for differences across speakers and items. We estimated the significance level of the fixed effect using Markov Chain Monte Carlo sampling [27].

| Fixed Effects | Est. | SE | t | pMCMC |
|---|---|---|---|---|
| Intercept | 2.045 | 0.17 | 12.06 | <.001 |
| Rhythmicity of couplet A | 0.16 | 0.07 | 2.35 | <.05 |

Table 1. *Parameter estimates of mixed-effects model predicting Rhythmicity of couplet B.*

| Fixed Effects | Est. | SE | t | pMCMC |
|---|---|---|---|---|
| Intercept | 1.45 | 0.18 | 8.08 | <.001 |
| Expressivity of couplet A | 0.43 | 0.06 | 7.61 | <.001 |

Table 2. *Parameter estimates of mixed-effects model predicting Expressivity of couplet B.*

Across speakers and verses, the Rhythmicity rating of the first couplet of a verse was predictive of the rating of the second couplet, *t* = 2.35, p < .05 (Table 1). The simple correlation between the Rhythmicity ratings of the first and second couplets are presented in Figure 3. This result demonstrates that, across speakers and items, the rhythmicity of the first couplet was predictive of that of the second couplet.

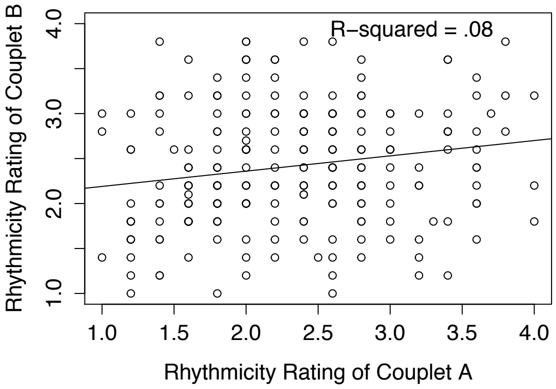


Figure 3: *Correlation of average perceived rhythmicity rating of the first and second couplet of verses.*


Figure 4: *Correlation of average perceived expressivity ratings of the first and second couplet of verses.*

Across speakers and verses, the Expressivity rating of the first couplet of a verse was predictive of the rating of the second couplet, *t* = 7.61, p < .001 (Table 2). The simple correlation between the Expressivity ratings of the first and second couplets is presented in Figure 4. This result demonstrates that, across speakers and items, the expressivity of the first couplet was predictive of that of the second couplet.

## 4. Discussion

In order to begin to explore the features of rhythmic children's books that support literacy development, we have constructed a spoken corpus of sixteen productions of *The Cat in the Hat* by Dr. Seuss. We gathered perceptual ratings of both the rhythmicity and expressivity of many of the verses from the text, and identified acoustic correlates of these perceptual ratings. Specifically, we found that rhythmicity ratings were correlated with variability in the timing of the onsets of the strong syllables of couplets and that expressivity ratings were correlated with F0 variability across words. These results demonstrate that naïve raters can successfully rate perceived expression and rhythm in speech [5].

We found no correlation between the rhythmicity and expressivity ratings of the individual couplets. This result demonstrates that expressive speech is not inherently any more or less rhythmic than non-expressive speech. Furthermore, it shows that participants are able to rate speech rhythmicity independently of speech expressivity, disproving an original concern of the study that participants would not be able to differentiate these two features.

Finally, we observed intraverse relationships for both the rhythmicity and expressivity ratings. That is, even after accounting for differences across individual speakers and individual verses, it is still the case that the rhythmicity or expressivity of the first couplet of a verse is significant predictor of the rhythmicity or expressivity of the second couplet. These results are particularly striking as the rhythmicity and expressivity of the first and second couplet of a verse were always rated by different sets of participants. This result suggests that readers of rhythmic texts are consistent in their productions of whole verses, suggesting that readers engage in self-entrainment in both their rhythmicity and expressivity.

It may be the case that, like the adults in the current study, children demonstrate self-entrainment when reading rhythmic prose, and that this process helps improve their rapid auditory processing, leading to improved reading fluency [13]. Follow-up studies will investigate (1) whether children reading *The Cat in the Hat* also show self-entrainment within verses and (2) whether reading overtly rhythmic prose improves children's overall fluency.

## 5. Conclusions

We have generated a corpus of child-directed speech that varies in its rhythmicity and expressivity. Moreover, we have identified acoustic correlates of these perceptual measures. We have demonstrated that there is no consistency between the rhythmicity and expressivity ratings of individual couplets. However, there is strong intraverse consistency for both rhythmicity and expressivity.

## 6. Acknowledgements

The authors would like to thank Suyin Taunton and Julia Karron for help with sound file alignment and Molly Morgan and Cam Vilain for assistance with running participants.

## 7. References

[1] Bhide, A., Power, A. & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education, 7,* 113-123.

[2] Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5,* 253-263.

[3] Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm**.** *Journal of Phonetics, 40,* 351-373.

[4] Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica 66*, 95-112.

[5] Dilley, L. & McAuley, J.D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59, 294-311.*

[6] Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41,* 233-243.

[7] Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465-494.

[8] Dunst, C.J., Meter, D., Hamby, D.W. (2011). Relationship between young children's nursery rhyme experiences and knowledge and phonological and print-related abilities. *CELL reviews, 4(1),* 1-12.

[9] Hayes, D. S. (2001). Young children's phonological sensitivity after exposure to a rhyming or nonrhyming story. *Journal of Genetic Psychology, 162,* 253-259.

[10] Stadler, M., Watson, M., Skahan, S (2007). Rhyming and vocabulary: Effects of lexical restructuring. *Communication Disorders Quarterly, 28(4),* 197-205.

[11] Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., et al. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proc. Natl. Acad. Sci. USA, 6,* 10911-10916.

[12] Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal Of Physiology, 102(1-3),* 120-129.

[13] Ritter, M., Colson, K. A., & Park, J. (2013). Reading Intervention Using Interactive Metronome in Children With Language and Reading Impairment: A Preliminary Investigation. *Communication Disorders Quarterly, 34(2),* 106-119.

[14] Port, R. F., Tajima, K., & Cummins, F. (1996). Self-entrainment in animal behavior and human speech. *Online proceedings of the 1996 Midwest artificial intelligence and cognitive science conference*, Indiana University, Bloomington, IN.

[15] Cummins, F. (2003): Rhythmic grouping in word lists: competing roles of syllables, words and stress feet. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain,* 325-328.

[16] Cummins, F. (2005): Interval timing in spoken lists of words. *Music Perception, 22,* 497-508.

[17] Dilley, L., Wallace, J., & Heffner, C. (2012). Perceptual isochrony and fluency in speech by normal talkers under varying task demands. *Prosodies: Context, Function, and Communication,* O. Niebuhr and H. Pfitzinger (Eds.), *Language, Context, and Cognition series,* Berlin/New York: Walter deGruyter, pp. 237-258.

[18] Payne, E., Post, B., Astruc, L., Prieto, P. & Vanrell, M. (2009). Rhythmic modification in child directed speech. *Oxford University Working Papers in Linguistics, Philology & Phonetics, 12,* 123–144.

[19] Nakata, T., & Trehub, S. E. (2011). Expressive timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicology: Music, Mind & Brain, 21,* 45-53.

[20] Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227–256.

[21] Fernald, A & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209-221.

[22] Thiessen, E., Hill, E., & Saffran, J. (2005) Infant-directed speech facilitates word segmentation, *Infancy, 7,* 5-71.

[23] Boersma, P. & Weenink, D. (2002). Praat, a system for doing phonetics by computer. Software and manual available online at: <http://www.praat.org>.

[24] Gorman, K., Howell, J. & Wagner, M. (2011). Prosodylab-Aligner: A tool for forced alignment of laboratory speech, Proceedings of Acoustics Week in Canada, Quebec City.

[25] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5 pp. 378–382.

[26] Baayen, R. H. (2008). Analyzing linguistic data: a practical introduction to statistics using R. Cambridge: Cambridge University Press.

[27] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.