### Abstract:
Temporal and phonological predictability in children's literature may support early literacy acquisition. Because such texts are typically heard before they are read, realization of predictive structure in caregiver prosody could guide children's attention during shared reading in a manner that supports reading subskill development. However, little is known about how predictive structure is realized in prosody during child-directed reading. We investigated whether speakers use word intensity to signal predictive metric and rhyme structure in a corpus of child-directed and read-along productions of The Cat in the Hat (Dr. Seuss, 1957). Using linear mixed-effects regression, we modeled the maximum intensity (dB) of each produced monosyllabic word as a function of metric strength, rhyme predictability, and a set of control parameters. In the control model, intensity increased with lower lexical frequency, capitalization, first mention, and syntactic boundary likelihood. Metric structure predicted word intensity beyond these control factors hierarchically, such that words were produced with highest intensity when aligned with beat one in a 6/8 metric structure, intermediate intensity when aligned with beat four, and lowest intensity when aligned with all other beats. Additionally, phonologically predictable rhyme targets were produced with reduced intensity. Together these results demonstrate that predictability is encoded in word intensity during child-directed reading, and that metric structure is realized hierarchically in word intensity. Further, the manner by which predictability is encoded in word intensity differs from that previously reported for word duration in this same corpus (Breen, 2018), demonstrating that intensity and duration present nonidentical prosodic information channels.
Metric structure and rhyme predictability modulate speech intensity during child-directed reading

Ahren B. Fitzroy\textsuperscript{1,2} & Mara Breen\textsuperscript{1}

\textsuperscript{1}Department of Psychology and Education, Mount Holyoke College

\textsuperscript{2}Department of Psychological and Brain Sciences, University of Massachusetts, Amherst

Corresponding author:

Ahren B. Fitzroy, Ph.D.

Mount Holyoke College
50 College Street
South Hadley, MA 01075
USA

E-mail: afitzroy@mtholyoke.edu
Temporal and phonological predictability in children’s literature may support early literacy acquisition. Because such texts are typically heard before they are read, realization of predictive structure in caregiver prosody could guide children’s attention during shared reading in a manner that supports reading subskill development. However, little is known about how predictive structure is realized in prosody during child-directed reading. We investigated whether speakers use word intensity to signal predictive metric and rhyme structure in a corpus of child-directed and read-alone productions of *The Cat in the Hat* (Dr. Seuss, 1957). Using linear mixed-effects regression, we modeled the maximum intensity (dB) of each produced monosyllabic word as a function of metric strength, rhyme predictability, and a set of control parameters. In the control model, intensity increased with lower lexical frequency, capitalization, first mention, and syntactic boundary likelihood. Metric structure predicted word intensity beyond these control factors hierarchically, such that words were produced with highest intensity when aligned with beat one in a 6/8 metric structure, intermediate intensity when aligned with beat four, and lowest intensity when aligned with all other beats. Additionally, phonologically predictable rhyme targets were produced with reduced intensity. Together these results demonstrate that predictability is encoded in word intensity during child-directed reading, and that metric structure is realized hierarchically in word intensity. Further, the manner by which predictability is encoded in word intensity differs from that previously reported for word duration in this same corpus (Breen, 2018), demonstrating that intensity and duration present nonidentical prosodic information channels.
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

Metric structure and rhyme predictability modulate speech intensity during child-directed reading

Two common features in children’s literature across languages are the presence of a strong metrical framework, and the use of rhyming phrases (Burling, 1966; Hanna, Lindner, & Dufter, 2002). The temporal and phonological predictability created by these features may support the purported benefits of these texts for literacy acquisition (Goswami, 1999; Huss, Verney, Fosker, Mead, & Goswami, 2011); early children’s literature is typically heard before it is read, and temporal and phonological predictability guide the allocation of attention to important moments in the auditory stream (Astheimer & Sanders, 2009, 2011; Breen, Dilley, McAuley, & Sanders, 2014; Fitzroy & Sanders, 2015). However, little is empirically known about how meter and rhyme in children’s literature are instantiated in spoken productions of these texts, which is a necessary precursor to understanding how such structures might guide perceptual or literacy learning in young listeners. Breen (2018) recently demonstrated that the metric and rhyme structures of The Cat in the Hat (Dr. Seuss, 1957) are realized in the duration and temporal spacing of words in child-directed productions of the text. Prominence in speech is imparted through several prosodic cues however, with duration and intensity typically thought to be the most important (Brenier, Cer, & Jurafsky, 2005; Kochanski, Grabe, Coleman, & Rosner, 2005; Kochanski & Orphanidou, 2008; Silipo & Greenberg, 2000). Therefore, in the present paper we examine the effects of meter and rhyme in The Cat in the Hat on the produced intensity of words during child-directed reading, and consider the results in light of previous findings of impacts on word duration (i.e., Breen, 2018).

Multiple studies have demonstrated that increased intensity imparts prominence in speech. Increased intensity for more prominent syllables has been observed for isolated and
continuous speech in which prominence is subjectively assessed by a listener (Brenier et al., 2005; Kochanski et al., 2005; Silipo & Greenberg, 2000; Streefkerk, Pols, & Bosch, 1999), and for metronome-synchronous speech in which prominence is objectively defined as syllabic alignment with a metronome click (Boutsen, Brutten, & Watts, 2000; Kochanski & Orphanidou, 2008). Intensity variation signals metric structure similarly during expressive musical performance, with greater intensity indicating greater metric strength. Importantly, intensity in expressive music performance imparts metric strength in a hierarchical manner: metric structures that involve three levels of strength (e.g., 6/8 meter) are produced with three corresponding levels of intensity (Drake & Palmer, 1993). It is unclear whether intensity in continuous speech imparts metric prominence in a similarly hierarchical manner, or only in a binary manner. The strongly hierarchical metric structure of The Cat in the Hat is reflected in hierarchically produced word duration (Breen, 2018), suggesting that it may also be signaled through hierarchical modulation of produced word intensity.

Predictability in speech also modulates the production of words, as more predictable words are typically acoustically reduced relative to less predictable words (Aylett & Turk, 2004; Gregory, Raymond, Bell, Fosler-lussier, & Jurafsky, 1999; Jurafsky, Bell, Gregory, & Raymond, 2001). Moreover, predictability-related reduction is preferentially realized in word intensity, such that more predictable words are produced with lower intensity regardless of whether they are repeated in the local context (Lam & Watson, 2010). In contrast, local word repetition is preferentially associated with shortened word duration (Lam & Watson, 2010). Most stanzas in The Cat in the Hat are written such that the first line ends with a rhyme prime (e.g., “all” in (1)), and the second line ends with a rhyme target (e.g., “fall” in (1)). This structure causes the stanza-final rhyme target to be both phonologically and semantically predictable, which should cause
the rhyme targets to be reduced during production. Breen (2018) observed equivalent word
duration but lengthened inter-word-onset intervals for predictable rhyme targets relative to
rhyme primes in *The Cat in the Hat*, which may be interpreted as a reduction of the rhyme target.
However, the temporal characteristics alone do not conclusively support this interpretation.
Given that predictability-related reduction is preferentially realized in word intensity (Lam &
Watson, 2010), we predict that word intensity will be reduced for rhyme targets relative to rhyme
primes in *The Cat in the Hat*.

In the current study, we test the hypotheses that metric structure and rhyme predictability
in *The Cat in the Hat* are realized in produced word intensity during child-directed reading. We
assess the influence of metric strength and rhyme predictability on word intensity in a corpus of
child-directed and read-alone productions of *The Cat in the Hat* (Breen, 2018), while controlling
for intrinsic word characteristics, font emphasis, and local repetition. We predict that metrically
stronger words will be produced with higher intensity, and that, similar to prior findings in
expressive musical performance, hierarchical metric structure will be realized as hierarchically
produced word intensity. We further predict that predictable rhyme targets will be produced with
reduced intensity relative to rhyme primes.

Method

Participants

Eighteen young adult (age 18-35 years) Mount Holyoke students participated in the
current study. All participants identified as female and as native speakers of American English,
defined as speaking English in the United States since at least age five. One participant was later
determined to not be a native speaker of American English and withdrawn from the study,
leaving 17 participants in the final analyses. All participants were compensated for their time with research participation credit for Psychology courses.

**Stimuli**

Participants read *The Cat in the Hat* (Dr. Seuss, 1957) aloud from a hardcover copy of the text. *The Cat in the Hat* is a 61-page illustrated book written primarily in anapestic (i.e., weak-weak-STRONG) tetrameter. It consists of 1625 words (1576 monosyllabic, 236 unique lexemes) organized primarily into seventy stanzas, with each stanza containing two lines of four anapests each (as in (1)). The first line in each stanza ends with a rhyme prime and the second line ends with a phonologically predictable rhyme target. In addition, there are six single anapestic lines in the text. In two cases, these single lines rhyme internally (e.g., And then something went BUMP! How that bump made us jump!). In the other four cases, the final word of the single line rhymes with the preceding stanza.

1. “Put me down!” said the fish.
   “This is no fun at all!
   Put me down!” said the fish.
   “I do NOT wish to fall!”

**Procedure**

Participants were randomly assigned to read *The Cat in the Hat* aloud in its entirety in one of two locations: to an audience of preschool students (4-5 years old) in a quiet classroom (n = 8), or alone in a quiet room (n = 9). In both environments participants read the book while seated and holding the book in their hands, and turned the pages on their own. All participants’ productions were digitally recorded using a head-mounted microphone (Shure SM10A) connected to a pre-amplifier (Rolls MP13 Mini-mic) plugged into the motherboard audio input.
of a laptop (in the classroom) or desktop computer (in the quiet room). Production recordings were digitized at 44,100 Hz with 16-bit resolution. Participants were given no specific instructions on how to read the text, other than to read it aloud. All participants provided informed consent before taking part in the study, and parents of the children who listened to the readings provided written assent for their child’s participation.

*Acoustic measures*

[FIGURE 1 HERE]

Figure 1 – Word intensity measurement. An excerpt from one *The Cat in The Hat* production is plotted as a time-frequency domain spectrogram (top) and as a time domain waveform (bottom). Identified word and silence boundaries are plotted as dashed vertical lines. The smoothed intensity contour generated for this excerpt is plotted in black over the spectrogram, with the parabolically-interpolated maximum intensity for each word indicated with an asterisk.

Word and silence boundaries (Figure 1) were identified in the production recordings by automatic force-alignment with the text in Praat (Boersma & Weenink, 2001) using the Prosodylab-Aligner (Gorman, Howell, & Wagner, 2011), which relies on the Hidden Markov Model Toolkit (Young & Young, 1993). Automatic force-alignment results were manually inspected and corrected as needed. As shown in Figure 1, linearly-spaced intensity (dB SPL) contours were calculated automatically in Praat for each production recording. Intensity contours were smoothed based on a minimum pitch of 100 Hz by squaring intensity values then convolving them with a 32 ms Gaussian window, which kept pitch-synchronous intensity variations to less than 0.00001 dB (Boersma & Weenink, 2001). Word intensity was defined as the parabolically-interpolated maximum of the intensity contour occurring within each word.
Multisyllabic words were excluded from further analyses, because the unstressed syllables in these words have reduced intensities for reasons unrelated to metric structure (Fry, 1955). Disfluent and incorrect word productions and were manually identified and excluded from further analyses, resulting in a loss of 473 out of 26,792 possible monosyllabic word productions (1.77%). The remaining maximum intensity values of correctly-produced monosyllabic words were centered and scaled to standard deviation units (i.e., Z-transformed) separately for each participant.

Data analysis

We analyzed the standardized maximum intensity of correctly-produced monosyllabic words using linear mixed-effects regression modeling implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) within R (R Core Team, 2015; RStudio Team, 2014). Data and analysis scripts can be downloaded from https://osf.io/xgy4t. We predicted maximum word intensity as a function of metric strength, rhyme predictability, linguistic control factors (number of phonemes, lexical frequency, word class, font emphasis, intra-stanza repetition, and syntactic boundary strength), and presence of a child audience.

[FIGURE 2 HERE]

Figure 2 – Regression predictors. Word intensity was modeled using linear mixed-effects regression with within-subjects factors of metric strength in a 6/8 metric structure (MS), rhyme predictability (RP), number of phonemes (#P), lexical frequency (LF), word class (WC), font...
emphasis (FE), intra-stanza repetition (ISR), and syntactic boundary strength (SBS). Lexical frequency values are rounded to the nearest tenth for clarity. See text for details.

Metric strength was defined by beat position in a 6/8 metric hierarchy (Figure 2), with beat 1 of each measure assigned metric strength 3 (strong), beat 4 assigned metric strength 2 (medium), and beats 2, 3, 5, and 6 assigned metric strength 1 (weak). The first stressed syllable in each stanza was considered beat 1 of the first measure of that stanza. This metric parsing is in agreement with previous linguistic metric descriptions of *The Cat in the Hat* as anapestic tetrameter in terms of stress placement (Nel, 2004), but adopts a musical perspective of metric phase in that the first stressed event in a phrase likely represents the start of a larger metric unit (i.e., a measure) with any preceding unstressed events considered anacrustic (Lerdahl & Jackendoff, 1983). Rhyme predictability was coded as 1 for verse-final rhyme target words (e.g., “fall” in (1)) and 0 for all other words.

To account for effects of basic word and text characteristics on produced intensity, several linguistic control factors were included in the regression models (Figure 2). The 1576 monosyllabic words in *The Cat in the Hat* were annotated using the Medical Research Council (MRC) Psycholinguistic Database (Coltheart, 1981) for number of phonemes (range = 1-6, $M = 2.54$, $SD = 0.75$) and lexical frequency as assessed by Kučera-Francis (K-F; Francis & Kučera, 1982) norms (log-transformed K-F frequency; range = 0-11.16, $M = 7.37$, $SD = 2.31$). For words that did not have raw K-F frequencies ($n = 18$), the K-F frequency of the singular or infinitival form of the word was substituted if available. There was no K-F frequency available for the word *plop*, so it was excluded from further analyses. The analyzed words were also annotated for syntactic word class (open [$n = 540$], closed [$n = 1035$]), font emphasis (normal font [$n = 1550$],
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

SMALL CAPS \( n = 25 \)) intra-stanza repetition (not repeated \( n = 1243 \), repeated \( n = 332 \)), and syntactic boundary strength as defined by the Left-Right Boundary model (LRB; Breen, Watson, & Gibson, 2011; Watson & Gibson, 2004) \( \text{range} = 0-5, M = 1.24, SD = 2.00 \)). For detailed description about how the LRB model was applied, please see (Breen, 2018); in brief, the LRB model assigns a higher probability of a phrase boundary with increases in the size of the material a speaker has just produced (the left-hand side) and increases in what they are going to produce (the right-hand side).

Our regression approach was to first create a fully saturated control model containing all linguistic control factors as both fixed effects and random slopes over participant, and audience as a fixed effect. Before being entered into the model as either a fixed effect or random slope, continuous predictors were centered and categorical predictors were recoded such that the sum of factor levels for each contrast was equal to zero. We then refined the control model by removing each fixed effect individually in order of ascending \( t \) magnitude, and using a likelihood ratio test to compare model fit with and without the fixed effect to justify its continued inclusion. The control model included only fixed effects that significantly increased model fit. We then added metric strength into the regression model as both a fixed effect and random slope over participant, coded using simple contrast coding with strength level 2 (medium; beat 4) as the reference level. This model did not converge, so we iteratively removed random slopes in order of least variance explained until the model converged. We then compared the resultant metric strength model to an updated version of the control model containing only those random slopes included in the metric strength model using a likelihood ratio test to determine whether metric strength predicted word intensity variance beyond the control parameters. Lastly, we added rhyme predictability to the model as a fixed effect and random slope over participant, and
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

compared the result to the prior model containing control factors and metric strength using a
likelihood ratio test to determine whether rhyme predictability accounted for unique word
intensity variance.

Results

Control model

The control model included number of phonemes, lexical frequency, word class, font
emphasis, intra-stanza repetition, and syntactic boundary strength as random slopes over
participant, and lexical frequency, font emphasis, intra-stanza repetition, and syntactic boundary
strength as fixed effects (Table 1). Audience presence did not have an effect on produced word
intensity ($p > 0.8$), and was not included in the final control model. Similarly, number of
phonemes and word class were not warranted for inclusion in the control model as fixed effects
($p’s > 0.3$). Across individuals, higher word intensity was predicted by lower lexical frequency
($\beta = -0.133, SE = 0.006, t = -22.542$), presentation in a capitalized font ($\beta = 0.784, SE = 0.065, t$
$= 12.101$), not being a repeat of recent material ($\beta = -0.164, SE = 0.022, t = -7.567$), and higher
likelihood of being a syntactic boundary ($\beta = 0.015, SE = 0.008, t = 1.888$).

<table>
<thead>
<tr>
<th>Random slopes</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0014615</td>
<td>0.03823</td>
</tr>
<tr>
<td>Number of phonemes</td>
<td>0.0022393</td>
<td>0.04732</td>
</tr>
<tr>
<td>Lexical frequency</td>
<td>0.0009228</td>
<td>0.03038</td>
</tr>
<tr>
<td>Word class</td>
<td>0.0076330</td>
<td>0.08737</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.0437573</td>
<td>0.20918</td>
</tr>
<tr>
<td>Intra-stanza repetition</td>
<td>0.0048553</td>
<td>0.06968</td>
</tr>
<tr>
<td>Syntactic boundary strength</td>
<td>0.0015851</td>
<td>0.03981</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$\beta$ estimate</th>
<th>$\beta$ standard error</th>
<th>$t$</th>
<th>$\chi^2 (p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.021509</td>
<td>0.0077745</td>
<td>2.777</td>
<td>-</td>
</tr>
<tr>
<td>Lexical frequency</td>
<td>-0.132559</td>
<td>0.005881</td>
<td>-22.542</td>
<td>58.354 (&lt; 0.001)</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.784229</td>
<td>0.064807</td>
<td>12.101</td>
<td>38.876 (&lt; 0.001)</td>
</tr>
<tr>
<td>Intra-stanza repetition</td>
<td>-0.163933</td>
<td>0.021664</td>
<td>-7.567</td>
<td>37.551 (&lt; 0.001)</td>
</tr>
</tbody>
</table>
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

Table 1 – Control model. Variance estimates and standard deviations are shown for each random slope over participant included in the control model. β estimates, standard errors, and t-values are shown for each fixed effect justified for inclusion in the control model, along with chi-square and p values from the likelihood ratio tests justifying its inclusion. Note that the random slopes over participant for number of phonemes, lexical frequency, intra-stanza repetition, and syntactic boundary strength were later removed from the control model to allow direct comparison with the experimental model containing metric strength.

Experimental models

[FIGURE 3 HERE]

Figure 3 – Fixed effect estimates (β) from final experimental model. Metric strength and rhyme predictability estimates are highlighted in black, linguistic control factors are shown in grey. Thicker horizontal bars indicate one standard error, thinner horizontal bars indicate two standard errors.

As shown in Table 2 and Figures 3 and 4, position in a 6/8 metric hierarchy predicted word intensity such that beat 1 was produced with highest intensity, beat 4 was produced with intermediate intensity, and beats 2, 3, 5, and 6 were produced with similarly low intensity (beat 1 vs. beat 4: $β = 0.322, SE = 0.058, t = 5.536$; beats 2, 3, 5, 6 vs. beat 4: $β = -0.191, SE = 0.054, t = -3.554$). Including metric strength in the regression model significantly improved model fit relative to an updated version of the control model containing the same non-meter factors as fixed effects and random slopes over participant, $χ^2 (11) = 1157.60, p < 0.001$. Therefore, metric strength provides explanatory power for produced word intensity beyond that explained by intrinsic and contextual word characteristics.
Table 2 – Initial experimental model containing metric strength. Variance estimates and standard deviations are shown for each random slope over participant included in the initial experimental model examining metric strength. β estimates, standard errors, and t-values are shown for each fixed effect included in this model.

<table>
<thead>
<tr>
<th>Random slopes</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.004285</td>
<td>0.06546</td>
</tr>
<tr>
<td>Beat 1 vs. beat 4</td>
<td>0.058568</td>
<td>0.24201</td>
</tr>
<tr>
<td>Beats 2, 3, 5, 6 vs. beat 4</td>
<td>0.041924</td>
<td>0.20475</td>
</tr>
<tr>
<td>Word class</td>
<td>0.020477</td>
<td>0.14310</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.051992</td>
<td>0.22802</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>β estimate</th>
<th>β standard error</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.138004</td>
<td>0.012368</td>
<td>11.158</td>
</tr>
<tr>
<td>Beat 1 vs. beat 4</td>
<td>0.322274</td>
<td>0.058215</td>
<td>5.536</td>
</tr>
<tr>
<td>Beats 2, 3, 5, 6 vs. beat 4</td>
<td>-0.191348</td>
<td>0.053841</td>
<td>-3.554</td>
</tr>
<tr>
<td>Lexical frequency</td>
<td>-0.115302</td>
<td>0.003829</td>
<td>-30.113</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.623397</td>
<td>0.071589</td>
<td>8.708</td>
</tr>
<tr>
<td>Intra-stanza repetition</td>
<td>-0.158835</td>
<td>0.014046</td>
<td>-11.308</td>
</tr>
<tr>
<td>Syntactic boundary strength</td>
<td>0.002515</td>
<td>0.003936</td>
<td>0.639</td>
</tr>
</tbody>
</table>

Figure 4 – Maximum word intensity as a function of position in a 6/8 metric hierarchy and rhyme predictability. Produced word intensity increased hierarchically with metric strength (left), and predictable rhyme targets were produced with lower intensity than other words, including phonologically similar but unpredictable rhyme primes (right). Note that all words that were not rhyme targets were coded as unpredictable in the regression models, but only unpredictable rhyme primes are shown here for clarity.

As shown in Table 3 and Figures 3 and 4, rhyme predictability predicted word intensity such that predictable rhyme targets (e.g., “fall” in (1)) were produced with lower intensity than other words, including metrically and phonologically similar rhyme primes ($\beta = -0.434, SE =$...
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

0.086, \( t = -5.030 \)). Adding rhyme predictability to the regression model significantly improved model fit relative to the model containing final control parameters and metric strength, \( \chi^2 (7) = 290.73, p < 0.001 \), demonstrating that rhyme predictability provides explanatory power for produced word intensity beyond that explained by intrinsic and contextual word characteristics and metric structure.

<table>
<thead>
<tr>
<th>Random slopes</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.003865</td>
<td>0.06217</td>
</tr>
<tr>
<td>Beat 1 vs. beat 4</td>
<td>0.035875</td>
<td>0.18941</td>
</tr>
<tr>
<td>Beats 2, 3, 5, 6 vs. beat 4</td>
<td>0.021121</td>
<td>0.14533</td>
</tr>
<tr>
<td>Rhyme predictability</td>
<td>0.111544</td>
<td>0.33398</td>
</tr>
<tr>
<td>Word class</td>
<td>0.022192</td>
<td>0.14897</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.058998</td>
<td>0.24290</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>( \beta ) estimate</th>
<th>( \beta ) standard error</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.175713</td>
<td>0.012311</td>
<td>14.272</td>
</tr>
<tr>
<td>Beat 1 vs. beat 4</td>
<td>0.204102</td>
<td>0.041646</td>
<td>4.901</td>
</tr>
<tr>
<td>Beats 2, 3, 5, 6 vs. beat 4</td>
<td>-0.302707</td>
<td>0.040018</td>
<td>-7.564</td>
</tr>
<tr>
<td>Rhyme predictability</td>
<td>-0.434422</td>
<td>0.086374</td>
<td>-5.030</td>
</tr>
<tr>
<td>Lexical frequency</td>
<td>-0.113800</td>
<td>0.003822</td>
<td>-29.776</td>
</tr>
<tr>
<td>Font emphasis</td>
<td>0.608770</td>
<td>0.074313</td>
<td>8.192</td>
</tr>
<tr>
<td>Intra-stanza repetition</td>
<td>-0.159876</td>
<td>0.013965</td>
<td>-11.448</td>
</tr>
<tr>
<td>Syntactic boundary strength</td>
<td>0.011017</td>
<td>0.003962</td>
<td>2.781</td>
</tr>
</tbody>
</table>

Table 3 – Final experimental model containing metric strength and rhyme predictability. Variance estimates and standard deviations are shown for each random slope over participant included in the final experimental model examining both metric strength and rhyme predictability. \( \beta \) estimates, standard errors, and \( t \)-values are shown for each fixed effect included in this model.

Discussion

As predicted, both metric structure and rhyme predictability in *The Cat in the Hat* modulated word intensity during read-aloud productions of the text. Greater metric strength resulted in greater produced word intensity in a hierarchic manner, such that the metrically
strongest words were produced with highest intensity, metrically intermediate words were produced with intermediate intensity, and metrically weak words were produced with lowest intensity. Rhyme predictability resulted in clear acoustic reduction, as rhyme targets were produced with lower intensity than other words including rhyme primes. Moreover, both metric strength and rhyme predictability provided unique explanatory power for word intensity, accounting for variance beyond that accounted for by lexical frequency, font emphasis, local repetition, and syntactic boundary strength. These effects were unique from one another as well, with rhyme predictability providing explanatory power beyond that accounted for by metric strength alone.

The effects of intrinsic word and text factors, local repetition, and syntactic structure in The Cat in the Hat on produced word intensity were largely consistent with prior findings for produced word duration (Breen, 2018). Consistent with the duration findings of Breen (2018), produced word intensity increased with lower lexical frequency, higher likelihood of being a syntactic boundary, and font emphasis in the form of SMALL CAPS, and decreased with repetition. Contrary to the duration findings of Breen (2018) however, produced word intensity was not modulated by word length, word class, or the presence or absence of a child audience. Together, these results demonstrate that the intrinsic and contextual linguistic information signaled by duration and intensity during child-directed reading is often redundant between these two prosodic channels, but not identical.

The realization of hierarchical metric structure in The Cat in the Hat as hierarchical word intensity demonstrates that during child-directed poetic reading, multiple levels of acoustic prominence are signaled. This regular, hierarchically organized prominence could provide temporal guideposts for young listeners, creating expectancies for certain moments in time to
which children can then predictively attend (e.g., Jones, 1976). Such periodic guidance of temporal attention during child-directed reading could represent a mechanism by which highly metrical children’s literature leads to positive reading outcomes; better ability to synchronize to an external auditory rhythm is associated with better pre-reading skills (e.g., phonological awareness and rapid automatized naming) in pre-readers (Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014), and guided practice tracking auditory rhythms during shared reading could improve auditory rhythm synchronization skills more broadly. Alternatively, increased intensity for metrically strong words could reflect speaker-centric factors such as greater speaker attention to metrically strong words (e.g., Arnold & Watson, 2015). The observed lack of audience effect could indicate that the metric strength effect is less likely to be driven by concerns about the listener. However, the read-alone participants were not given any specific instructions on how to read the text, and we speculate that because the experimental material is a prominent children’s book, the read-alone participants may have read it as they would have to a child. Importantly, whether the metric strength effect reflects speaker-centric or listener-centric motivations, hierarchical prominence in the speech signal could still provide perceptual benefits to the listener.

The realization of hierarchical meter in *The Cat in the Hat* as hierarchical word intensity is consistent with previous report of its realization as hierarchical word duration and inter-word intervals (Breen, 2018). However, the specific pattern of metric prominence realized in intensity differs from that realized in temporal variation. In the present study, word intensity was highest for beat 1 in a 6/8 metric parsing, intermediate for beat 4, and lowest for beats 2, 3, 5, and 6. Conversely in Breen (2018), word duration and inter-word intervals were longest for beat 4 in a 6/8 parsing (called metric levels 3, 4, and 5 in the metric model inspired by Fabb & Halle (2008).
employed by Breen), intermediate for beat 1 (called metric level 2), and shortest for beats 2, 3, 5, and 6 (called metric level 1). These findings demonstrate that although word intensity and word duration provide important and potentially overlapping cues to prominence in speech, the information encoded on these two prosodic channels is not identical. In the same productions, word duration was aligned with the predictions of a linguistic model of rhythm in poetry (Fabb & Halle, 2008), whereas word intensity was aligned with previous observations of dynamics in expressive music performance (Drake & Palmer, 1993). Though it is possible that the observed dissociation between intensity and duration is unique to productions of poetry, which involves elements of linguistic and musical rhythm, previous reports that prominence in non-poetic contexts is best explained by a combination of intensity and duration (Kochanski & Orphanidou, 2008; Silipo & Greenberg, 2000) suggest that this dissociation may be a more general prosodic mechanism. Further, these findings indicate that metric strength is realized in both word duration and word intensity, but that consistent with prior work (Wagner & Watson, 2010) phrase structure is preferentially encoded in word duration; in *The Cat in the Hat*, beat 4 in a 6/8 metric parsing often occurs at phrase-final positions, whereas beat 1 never does. Though we did not explicitly model for phrase structure in either study, our results across the two studies could be interpreted as hierarchically increased duration and intensity with metric strength in a 6/8 structure, combined with an additional duration increase for phrase-final beat 4s.

The observed intensity reduction for highly predictable rhyme targets is consistent with prior findings that more predictable words are acoustically reduced relative to less predictable words (e.g., Jurafsky et al., 2001). Moreover, the contrast between the observed clear predictability-related reduction in intensity and the equivocal predictability effects on duration presented by Breen (2018) is consistent with prior findings indicating that predictability-related...
reduction is preferentially realized in intensity (Lam & Watson, 2010). There is no local perceptual benefit to a listener when a talker acoustically reduces a predictable word, as reduction reduces the effective signal-to-noise ratio of that speech element. Further, listeners do not preferentially allocate attention to predicted moments in speech when the information to be presented at that time is also completely predictable, reducing the signal to noise ratio even more at these times (Astheimer & Sanders, 2011). Nonetheless, highly predictable words are regularly communicated without error, suggesting that the system is robust to information loss at moments where information is highly predictable. It may be that listeners shift from a detailed, attentive perceptual strategy to a template-matching perceptual strategy when information is highly predicted, which would be both more efficient and more robust to reduced signal-to-noise ratio.

Note that reductions in produced word intensity might be a result of either speakers being aware of and accommodating such a strategy by listeners or of a similar shift in strategy by the speaker during production without regard to the listener, but in both cases the communicative outcome is the same.

Collectively, the increased prominence with metric strength and reduction with rhyme predictability observed in the current study represent two methods by which predictability modulates produced word intensity. One one hand, temporal predictability, such as that provided by a regular metric framework, indicates when important information is likely to occur but not what that information will be. It is advantageous then for listeners to direct attention to these moments, maximizing the perceptual resources available to encode the unknown important information. It is in turn advantageous for talkers to impart prominence at these moments, both to attract listener attention and to increase signal strength. On the other hand, phonological and semantic predictability provide strong expectations regarding not only when important
RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY

Information will occur, but also what that information will be. Under these conditions it is less important for listeners to encode with high perceptual detail, instead only needing to encode with sufficient detail to confirm or disconfirm their expectations. In turn, talkers can take advantage of the lowered communicative demands of highly predictable information by reducing production effort for such words. Finally, the differences between how metric strength and rhyme predictability are realized in word intensity in the present study compared to in word and inter-word duration in Breen (2018) provide further evidence that intensity and duration provide important, but separate, prosodic communication channels during speech production.

Funding

This work was supported by the James S. McDonnell Foundation [Understanding Human Cognition Scholar Award to MB].
References


RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY


RUNNING HEAD: METER AND RHYME MODULATE SPEECH INTENSITY


Figure 1 – Word intensity measurement. An excerpt from one The Cat in The Hat production is plotted as a time-frequency domain spectrogram (top) and as a time domain waveform (bottom). Identified word and silence boundaries are plotted as dashed vertical lines. The smoothed intensity contour generated for this excerpt is plotted in black over the spectrogram, with the parabolically-interpolated maximum intensity for each word indicated with an asterisk.

86x41mm (300 x 300 DPI)
Figure 2 – Regression predictors. Word intensity was modeled using linear mixed-effects regression with within-subjects factors of metric strength in a 6/8 metric structure (MS), rhyme predictability (RP), number of phonemes (#P), lexical frequency (LF), word class (WC), font emphasis (FE), intra-stanza repetition (ISR), and syntactic boundary strength (SBS). Lexical frequency values are rounded to the nearest tenth for clarity. See text for details.
Figure 3 – Fixed effect estimates (β) from final experimental model. Metric strength and rhyme predictability estimates are highlighted in black, linguistic control factors are shown in grey. Thicker horizontal bars indicate one standard error, thinner horizontal bars indicate two standard errors.
Figure 4 – Maximum word intensity as a function of position in a 6/8 metric hierarchy and rhyme predictability. Produced word intensity increased hierarchically with metric strength (left), and predictable rhyme targets were produced with lower intensity than other words, including phonologically similar but unpredictable rhyme primes (right). Note that all words that were not rhyme targets were coded as unpredictable in the regression models, but only unpredictable rhyme primes are shown here for clarity.