Journal of Eye Movement Research
13(3):5

Rhythmic subvocalization: An eye-tracking study on silent poetry reading

Judith Beck
Cognitive Science
University of Freiburg, Germany

Lars Konieczny
Cognitive Science
University of Freiburg, Germany

The present study investigates effects of conventionally metered and rhymed poetry on eye-movements in silent reading. Readers saw MRRL poems (i.e., metrically regular, rhymed language) in two layouts. In poem layout, verse endings coincided with line breaks. In prose layout verse endings could be mid-line. We also added metrical and rhyme anomalies. We hypothesized that silently reading MRRL results in building up auditive expectations that are based on a rhythmic "audible gestalt" and propose that rhythmicity is generated through subvocalization. Our results revealed that readers were sensitive to rhythmic-gestalt-anomalies but showed differential effects in poem and prose layouts. Metrical anomalies in particular resulted in robust reading disruptions across a variety of eye-movement measures in the poem layout and caused re-reading of the local context. Rhyme anomalies elicited stronger effects in prose layout and resulted in systematic re-reading of pre-rhymes. The presence or absence of rhythmic-gestalt-anomalies, as well as the layout manipulation, also affected reading in general. Effects of syllable number indicated a high degree of subvocalization. The overall pattern of results suggests that eye-movements reflect, and are closely aligned with, the rhythmic subvocalization of MRRL.
This study introduces a two-stage approach to the analysis of long MRRL stimuli and contributes to the discussion of how the processing of rhythm in music and speech may overlap.

Keywords: Eye movements, poetry, silent reading, rhythm, subvocalization, auditive expectation, meter, rhyme, load contribution

To recapitulate then:
I would define, in brief,
the Poetry of words as
The Rhythmical Creation
of Beauty.

Edgar Allan Poe, The Poetic Principle

Introduction

If you know Grimm's fairy tale Rumpelstiltskin, you might agree that these lines are funny: *Ha! glad am I that no one knew, | that rhythmic guitar I do play well.*

Why is that? The original – *"Ha! glad am I that no one knew | That Rumpelstiltskin I am styled"* (Grimm, 1884) – has a strict metrical structure. This creates a regular rhythmic pattern, whereas in the introductory example, the stress pattern deviates from the expected metrical scheme. It is difficult to accommodate to that deviation rhythmically. Thus, it prompts an experience of rhythmical oddness in the second line (which you might have just experienced while reading silently). That, in turn, justifies the sentence's content, i.e., why it might be good that nobody knows that the speaker plays rhythm guitar, as he or she appears to be lacking talent.

While establishing speech rhythm is not at all trivial, rhythmic patterns appear to be a relatively easy cognitive task in regularly metered and rhymed poems even for children (Rubin et al., 1997). However, it remains unclear how it works in silent reading MRRL (i.e., metrically-regular, rhymed language), if *subvocalization* plays an important role in it, and whether eye-movements may reflect that.

Received November 24, 2019; Published September 14, 2021.
Citation: Beck, J. & Konieczny, L. (2021). Rhythmic subvocalization: An eye-tracking study on silent poetry reading. *Journal of Eye Movement Research*, *13*(3):5.
Digital Object Identifier: 10.16910/jemr.13.3.5
ISSN: 1995-8692

Subvocalization and eye-movements

Subvocalization serves to ‘prepare for pronunciation’, but is simultaneously characterized by inhibited speech-motor articulation (Laubrock & Kliegl, 2015, p. 2). It may involve hearing an *inner voice* (Abramson & Goldinger, 1997; Chafe, 1988; Huey, E., 1908; Perrone-Bertolotti et al., 2014). Fluent silent reading is usually preceded by the different stages of learning to read orally (Manguel, 1996), supposedly with varying degrees of subvocalization. Thereby phonological awareness is crucial for expressing prosody as well as coordinating temporal predictions for speech rhythm (Melby-Lervåg et al., 2012). When reading silently, a reader’s inner voice can also distinguish inertly between various qualities, mirroring external intonation and modulation, varying i.a., by volume/stress, pitch and tempo (Vilhauer, 2017). Hence, subvocalization can affect silent reading (Kriukova & Mani, 2016; Stolterfoht et al., 2007), which is reflected in the implicit prosody hypothesis. It states that phonological features influence syntactic parsing and guide ambiguity resolution (Bader, 1998; Fodor, 1998, 2002a, 2002b). For example, gaze durations and fixations are modulated by the number of stressed syllables within a word (Ashby & Clifton, 2005). Also, syntactic analysis is affected by the *alternating* distribution of stressed and unstressed syllables (Kentner, 2012, 2016; Kentner & Vasishth, 2016) and stress perception can be influenced by suprasegmental cues such as the preceding stress distribution (Brown et al., 2015). Importantly, in experiments using Limericks, syntactic reanalysis of a critical region elicits longer reading times when it requires a reanalysis of the metrical pattern (Breen & Clifton, 2011, 2013). This supports the notion of stress expectation management (Schmidt-Kassow & Kotz, 2009a, 2009b). These findings indicate that (lexical) stress registered in eye movements is based on phonological mental representations and that “readers form an implicit metrical representation of a text during silent reading” (Breen & Clifton, 2013, p. 1896). ERP results presented by Breen et al. (2019) offer first evidence that explicit and implicit metric may be processed similarly.

However, to date, the nature of this representation is not yet well understood (Breen, 2014). Is it abstract in the sense that it is a-modal, i.e., stripped of any sensory-motoric representations, and does it require at least some representation of sound, or even the - yet suppressed - execution of motor-action? There is initial evidence that eye movements during silent reading are influenced by subvocalization (Eiter & Inhoff, 2010) to the point that the spatial distance between the eyes (that lead) and the voice (that follows) might affect and even regulate eye movements (Laubrock & Kliegl, 2015), even in silent reading. With this in mind, subvocalization should play a key role in silent reading for the adaption of a MRRL-text’s metrical figures and rhythmic contour. The following questions arise: 1. Would an MRRL-rhythm, bearing a ‘purposeful’ audible ‘gestalt’, be perceivable to readers reading *silently*? 2. If so, would eye movements show sensitivity to anomalies within a rhythmic ‘gestalt’? 3. Can we observe eye-movements suggesting *subvocalization* of a rhythmic gestalt and if so, in which measures (Rayner, 2009; Rayner & Pollatsek, 1989)?

Metrically regular, rhymed language (MRRL)

Poetry, with *traditional* meter and rhyme, is considered melodic, being both music *and* language (Menninghaus et al., 2018). Historically, oral traditions preceded written compositions. Furthermore, a major aspect is rhythmicity. For Poe, the ‘rhythmical creation’ is essential for the ‘poetry of words’, i.e., a poem’s rhythm is caused by words' respective sounds appearing in a specific order, by which they build an audibly perceivable ‘gestalt’ (Carper & Attridge, 2003; Koelsch & Siebel, 2005; Lerdahl, 2001; Lerdahl, F., 2013; Metz-Göckel, 2008; Morgan et al., 2019; Slana et al., 2016; Tsur et al., 1991). In silent reading MRRL, this ‘gestalt’ would then have to be instantiated by the reader.

a) *Meter* is a contributing factor to an audible gestalt (Falk et al., 2014) in MRRL. In oral reading, its hierarchical nature is realized via two distinct means, intensity (Fitzroy & Breen, 2020) and duration (Breen, 2018). Traditionally, it relates to the percept of an alternation of stressed/accented (strong) vs. unstressed/unaccented (weak) syllables (Obermeier et al., 2013; Port, 2003; Selkirk, 1986). Meter is proposed to influence cognitive fluency, memory and verbatim recall (compare Andreetta et al., 2021; Obermeier et al., 2016; Tillmann & Jay Dowling, 2007; Van Peer, 1990), and to aid temporal-based predictive language processing and comprehension (Essens & Povel, 1985; Menninghaus et al., 2017; Rothermich et al., 2012).

We define meter according to Ravignani & Madison (2017) as the “hierarchical organization of temporal events based on stress and other spectral properties, such as loudness alternation, pitch variation, etc.” While *silently* reading MRRL, hierarchically, temporally, and spectrally shaped stress patterns may be represented, inferred and automatically categorized into metrical entities (i.e. a (linguistic) metrical grid, see Lerdahl, 2001, p. 5). An abstraction of their overall sonic similarity distribution is then projected onto what is expected in the next line or stanza, altering the "mode of attention" (Gjerdingen, 1989).

The frequent and structured repetition of a set of metrical figures contributes to their prominence and perception as regular. This, in turn, allows for beat extraction and induction (Honing, 2012, 2019). Here, beat is "psychologically superimposed" and can be defined as the "isochronic grid generated via metrical expectations" (Ravignani et al., 2019; Ravignani & Madison, 2017, p. 2). This grid is marked by a "rhythmic pattern, where all intervals have *roughly* equal duration" (ibid., authors emphasis), whereby rhythm can be understood as a durational-bound pattern of events within a time-frame (ibid., see also Lerdahl & Jackendoff, 1983; Schofield, 2016; Wade, 2004; Whittall, 2011). Due to its phonological rhythmicity, durational pattering, and structural repetition of meter *MRRL* supposedly offers a higher level of 'isochronicity' than normal speech (for an investigation of stable periodicity see Ravignani & Madison, 2017).

Although an inferred beat may inwardly 'go on' autonomously while reading, the respective prominent metrical figure has to be checked and updated in order to maintain it. Therefore, it must be aligned constantly with the upcoming input (for normal speech see Beier & Ferreira, 2018). This process is based on two levels: a) downright processing of local stress grids as required by the phonemic-syllabic material (Lerdahl, F., 2013, p. 261) and b) actualizing the underlying (physically or non-physically salient) quasi-isochronic MRRL-beat. However, both, rhythm and meter can change from one stanza to another or even from one verse to another. In this case, readers must attune their temporal predictions, either by inferring, respectively projecting a new 'metrical grid' to the following lines, or by adjusting to an accelerated/slowed-down beat, i.e., applying slightly increased or decreased intervals (Ravignani & Madison, 2017). If reading MRRL silently not only demands beat extraction but also requires successive beat induction (Honing, 2018) it should, in turn, create tension, and, if an expectation is not fulfilled, a sense of violation.

Possible changes to a rhythmic structure can be a violation of the number of (inferred) beats, a deviation by one syllable less or more, a substitution of a word by another one which demands preponed or delayed accent/stressing (see Arnal et al., 2015 for rhythmic tone sequences), or a new "sound" (vowel) that changes a gesture. Therefore, characteristics of phonemes, such as tonal weight/sonority, duration level, loudness, breathiness, e.g. /s/ vs. /a/, are likely to contribute to recognition and processing in silent MRRL reading at a very early stage, as it does in oral speech (Schmidtke et al., 2014; Yoncheva et al., 2013). These contrastive and coordinative features (Nolan & Jeon, 2014) as well as their related articulatory, co-articulatory and accentual gesture qualities (Tilsen, 2019), creating i.a. phenomena such as sound diffraction or floating stress, are at play in the consecutive order of syllables/words. Without this order, there would be stress but no 'regular meter', no 'beat' – and no *structured* MRRL-rhythm. Therefore, we propose that for MRRL, like for music, "meter involves *when* events will happen, while grouping involves *what* events will happen" (London, 2012b, p. 6). This notion is based on the assumption that for MRRL, "the strongest correspondences between music and language appear to be between musical syntax and linguistic phonology, not musical syntax and linguistic syntax" (Lerdahl, F., 2013, p. 257).

b) *Rhyme* as a stylistic device is the second important factor shaping the 'audible gestalt'. It contributes to a text's coordinating auditive characteristics by structuring the stream of words, respectively syllables, via repetition (Fabb, 2015) and via sonic modification, e.g. perfect vs. imperfect rhymes (Knoop et al., 2019; Schrott & Jacobs, 2011, p. 350), such as 'blind/mind' vs. 'line/find'. Readers or listeners of a poem seem to be sensitive towards rhyme schemes (Carminati et al., 2006; Obermeier et al., 2016). Scheepers et al. (2013), for instance, found strong effects of rhyme anomalies in listeners' pupillary responses. Hence, in MRRL, the formation of expectations of what is to be 'heard' or 'seen' in the next line is also triggered by the circulation of end rhymes as part of a larger time scale (Fabb, 2009) or internal rhymes (Hurschler et al., 2015; Kayser, 2002), supposedly related to smaller time scales. Importantly, as suggested by Schrott & Jacobs (2011, p. 352), the verse-end position of rhyme is crucial for determining the meter of a line, and at the same time divides a poem into segments. The poem's line as the salient and fundamental structural unit of verse marks boundaries for readers, in conventional poetry mostly via end-rhymes (Fabb et al., 2008; Fechino et al., 2020; but see Hetherington & Atherton, 2020 for the genre of prose poetry).

These boundaries often elicit pausing, supposedly due to closure effects (Smith, 1968) and enhanced by the poem's visual presentation. Naturally, pauses are crucial, too, for the production and detection of MRRL-rhythm. They have an explicit attentional function for maintaining rhythm (Fuller, 2001) and for directing breathing patterns in oral reciting, bearing the potential to be mapped onto subvocalized reading patterns. Turner & Pöppel (1983) proposed a time unit per verse (2-4s, average peak around 2.5-3.5s), which may shape reading/reciting MRRL and contribute to temporal prediction and segmenting, regardless of number of syllables (for a critique of the 3-sec-postulation see Fabb, 2013; but see Kien, J. & Kemp., A., 1994 for a comparison of durations of lines with biological action units; Wang et al., 2015, 2016; for a review see Yu & Bao, 2020). Ultimately, however, the overall regularity which allows for MRRL-rhythm because of stylistic devices, such as meter and rhyme or other parallelistic dictions (for details see Menninghaus et al., 2017; and Menninghaus & Blohm, 2020), is dependent on the phonological material of a poem (Kiparsky, P., 2009).

c) *Layout*. Most of the ongoing discussion about the role of layout, i.e., poem vs. prose (Fabb, 2009; Hanauer, 1996), as well as the function of features such as rhyme, focuses on the potential to affect categorization, reading strategy and tempo, comprehension and memory processes as well as aesthetic appreciation (Hanauer, 1998a, 1998b; Hoffstaedter, 1987; Menninghaus & Wallot, 2021; Peskin, 2007; Xue et al., 2020; Zwaan, 1991). Important in the context of our study is that when reading poetry, top-down processes, termed genre-effect, can impact attention strategies (Hanauer, 1996, 1998b). Furthermore, eye-movement patterns can differ when the same text is presented in poetry or in prose layout. Fechino et al. (2020) found overall longer gaze durations and a higher rereading probability in the poetry layout. Our present study has a similar design but different focus, and was completed and submitted before Fechino et al.'s study was published. The same holds for findings on the processing of rhyme and meter (Menninghaus & Wallot, 2021). Amongst other results, they report 'total gaze durations' to be longer for verse-final words, when either rhyme or meter or both were present, and findings were interpreted within the aesthetic emotions approach (but see Skov & Nadal, 2020).

Entrainment and MRRL

As stated earlier, to perceive a rhythm or to induce a beat, we must be able to synchronize and/or to entrain to a stimulus (Honing, 2012). Importantly, the term entrainment originally refers to *external* stimuli provoking an internal pattern of neuronal responses that seem to be rhythmically aligned with and periodically reflect or represent (external) stimuli, such as light (Floessner & Hut, 2017, p. 48), sound (Fujioka et al., 2012), rhythmic auditory stimuli (Nozaradan, 2014) or music (Tierney & Kraus, 2013, 2015). With speech, processing appears to be bound to timing patterns, too, for the auditive input as well as for the responding neuronal activation (Zoefel et al., 2018). Kotz et al. (2018, p. 896) propose that "we seem to neurally synchronize with rhythm in speech, which captures our attention, regularizes speech flow, [and] may emphasize meaning" (Kotz & Schwartze, 2016). In silently reading MRRL, perceived meter and rhythm may affect neurocognitive oscillators (Port, 2003) and may elicit a perception of periodicity (Kotz et al., 2018), even in the absence of an explicit signal. This, in turn, may lead to synchronization with an isochronal pulse (but, in terms of music, may not, see London, 2012a). Further support comes from the fact that production and processing of music and language share neuronal circuits (Fedorenko et al., 2009; Kunert et al., 2015; Patel, 2010; Rebuschat et al., 2011). As entrainment to music goes along with beat induction (Honing, 2012), we presume that beat extraction and induction works for MRRL, too, and may be a theoretical basis for the explanation of the cognitive phenomenon of rhythm effects (Obleser & Kayser, 2019, p. 913).

Aim and rationale of the study

To our knowledge, up until now no one has investigated the role of subvocalization linked to rhythm in silent reading of MRRL-poetry. Here, we propose that MRRL serves as an acoustic stimulus inwardly brought to mind via rhythmic subvocalization. As such, it is bound to timing and bears the potential to be entrained (Di Liberto et al., 2015; Kösem et al., 2018; Kotz et al., 2018, p. 902; Kotz & Schwartze, 2010; Merker et al., 2009; Tierney & Kraus, 2015). Accordingly, we hypothesize that readers pick up MRRL-rhythm when they read with an inner voice and, thus, that they should experience a sense of violation if the accuracy and predictability of MRRL is interrupted. The question is if and how this is reflected in eye movements.

Beyond phonological properties of MRRL, we were interested in the extent to which the line layout contributes to the rhythmic perception of MRRL-poems. If line breaks are used as additional rhythmic cues, it should, on the one hand, be more difficult to pick up the metrical grid and rhythm structure when poems are presented in prose form, i.e., when verse endings do not always coincide with actual line breaks. On the other hand, because the rhythmic and audible 'gestalt' of MRRL must be updated constantly, we suspect that reading is influenced by a text's (poem/stanza) sonic cues (compare Aryani et al., 2016; for a general discussion of the importance of phonologicy see Berent, 2013). Hence, MRRL rhythm should also be picked up in the prose layout, albeit leading to different eye movements compared with the poem layout.

So, firstly, we were interested in whether readers would take on an MRRL-rhythm at all. To test this, we introduced three types of anomalies at significant places in the poems: *metric* anomaly, *rhyme* anomaly, and a *combination* of both. A metric anomaly is a deviation of the expected linguistic metrical grid, at a specific location. This grid should govern the subvocalization of the line/stanza until the rhythmic inconsistency has to be processed. Salient deviations should result in a noticeable slowing-down in reading if they are experienced as 'violations' (compare Breen & Clifton, 2013), which would imply that the MRRL-rhythm had been picked up. For rhyme anomalies, Scheepers et al. (2013) report stronger reactions in pupil dilation than for metric or other anomalies. Thus, we also would expect rhyme anomalies to elicit longer reading times. For combined rhyme and meter anomalies, the single effects for rhyme and meter could, on the one hand, add up and thus lead to the longest reading times for this anomaly type. Also, the combined anomaly might impede the accommodation of the rhyme scheme. However, on the other hand, the combination could lead to the disintegration of the rhythmic structure, i.e., this anomaly might not be experienced as an expectation violation at all.

We expected the type of anomaly to interact with the line-layout of the poem. In the *poem* layout, the original verse-structure is preserved, whereas in the *prose* layout, line breaks, most of the time, do not coincide with verse endings. In the *poem* layout, the rhyme structure is clearly identifiable, as the end of verses coincide with line endings, whereas the rhyme words are hidden somewhere within the lines in the *prose* layout. This might have two consequences: First, it should be harder to pick up the rhyme scheme in the *prose* layout, and hence divergences from the given rhyme scheme might go unnoticed. Secondly, if a rhyme anomaly is detected, the pattern of refixations might differ, as the first word of a rhyme pair – called pre-rhyme (Smith, 1968) throughout the rest of the paper – is harder to detect in the prose layout, as its position is presumably more difficult to memorize.

The layout might also affect the processing of metric anomalies, because their detection might be easier in layouts with a strict verse-by-verse structure typical for poems.

Furthermore, we also expected re-fixations to the origin of the anomalies where possible. In rhyme anomalies, the origin is the corresponding rhyme word usually at the end of a verse above, whereas meter has no such clear origin, as it is construed across entire verses. However, since the units of rhythmic gestalt are comprised of only a few syllables, the immediate context of a metric violation is much more important than for rhyme anomalies. This should result in more local re-fixation patterns for metric violations, regardless of layout. Rhyme anomalies are expected to elicit more across-line refixation on the pre-rhyme.

On a more general level, we were also interested in identifying indicators of MRRL-triggered subvocalization in our eye-tracking parameters throughout entire poems. In particular, we were interested in how the introduction of anomalies and the layout versions would modulate reading in general, not only at critical interest areas (see *figure 1*).

Note that we have not included obvious semantic or syntactic anomalies in this study. Poems (or poetic language more generally) may induce a certain tolerance towards these kinds of violations (see Blohm et al., 2017 for investigation of genre-related tolerance towards semantic and morphological anomalies in verse, and syntactic inversions, 2018), but this is not a research question of this paper.

	MRRL_version (consistent/inconsistent)	*anomaly_type (rhyme, meter, rhyme+meter)*	*layout (poem/prose)*	*Number of syllables/CVQ*
Critical IAs (main model)	If readers subvocalize rhythmically, anomalies/inconsistent versions should elicit longer reading times plus more refixations on previous material.	For factor anomaly type, we only expect effects of inconsistent versions, i.e., an interaction with MRRL_version *Metric* anomalies should elicit longer reading times. *Rhyme* anomalies should also elicit longer reading times. Combined *rhyme&meter* anomalies should show the strongest effect, if the simple effects sum up, or because the accommodation of the rhyme scheme is impeded. Combined anomalies might also be too strong a divergence and might not be experienced as an anomaly at all.	If the visual layout functions as a structural cue, rhythmic subvocalization should be more pronounced in poem layout. Hence the above effects should vary in strength. Also, refixating on the pre-rhyme should be particularly hard in prose layout.	(variable not included in model)
Non-critical IAs (complete model)	a) Inconsistencies result in more cautious, slower reading b) Inconsistencies result in less rhythmic, faster reading	not included in complete model	If the visual layout functions as a structural cue, rhythmic subvocalization should be more pronounced in poem layout and should also lead to overall longer reading times compared to prose. Also, other variables that reflect subvocalization should show stronger effects in layout poem.	Number of syllables & Consonant vowel quotient (cvq) were included as predictors for reading times on top of word length (and other lexical variables) because they are presumably directly linked to pronunciation and thus indicators for subvocalization. We thus expected reliable effects of both variables, and potentially interactions with variables that modulate subvocalization.

Figure 1. Illustration of hypotheses for main and complete model

Methods

Participants

Thirty-eight participants (23 females; 15 males, mean age: 28.87 years; SD age = 12.33 range: 19-76 years) took part in the study. They were recruited via the Sona System within the Department of Psychology, University of Freiburg, Germany, via notices on bulletin boards at different faculties and via email to distributors like art associations, the House of Literature Freiburg or the Freiburg University of Music. All participants were native speakers of German. 29 were students or participants without clear indication of profession/status (average age: 24.9), 8 were employees (average age: 37.38), 1 was retired (aged 76). All of them had normal or corrected-to-normal vision and were naïve to the experiment's purpose. Subjects received either course credit for participating or alternatively signed up for a lottery drawing with the chance to win 3x 45 min of scientific or creative writing training.

Ethical Statement

No invasive or unsafe methods were applied and only behavioral data such as eye tracking data and questionnaires were collected. All participants gave written consent before the experiment started. The experiment was conducted in accordance with the standards of the "Ethical Principles for Medical Research Involving Human Subjects" (Declaration of Helsinki, 1964), set by the World Medical Association. This study was conducted according to the DFG-guidelines for good scientific practice, including originality of research idea, experimental design and method used, and is devoted to fair research behavior.

Apparatus

The experiment was designed and the study was conducted in spring/summer 2019 in the Cognitive Science eye-tracking laboratory at the *Center for Cognitive Science*, University of Freiburg. The reading experiment was set up with the 'Experiment builder' software (SR Research Ltd., Mississauga, Canada). Using the SR Research EyeLink 1000 (SR Research Ltd.) eye-tracking system, participants' eye movements were recorded, with a sampling rate of 500 Hz and an accuracy of 0.25° to 0.5° of the visual field. To reduce head and body movements, a chin-and-head rest was securely mounted on a table. The distance between the EyeLink 1000 chin-and-head rest and the screen was 60 cm. Only the right eye was tracked. Before eye-movement recording was started, standard 9-point calibration and validation procedure was executed to gain a spatial resolution error of less than 0.5° of the visual angle.

Design and Materials

Stimuli consist of eight German poems that were each manipulated according to a 2x2 design, comprising the factors *layout* (poem vs. prose) and *version* (original poems vs. versions that included rhythm and rhyme violations). The 8x4 (items x condition) texts were then distributed to four presentation lists following a Latin square rotation scheme, such that each participant was presented with two texts for each condition, and each item occurred only once per list.

The order of presentation of stimuli was randomized. Stimuli were presented in Trebuchet MS, with a font size of 30. The display resolution was 1920 (width) x 1080 (height) pixels, leaving space for up to 13 lines of text with a 1.5 line spacing. Stimuli were split over max. 3 pages of the screen (for poem versions: page one presented stanza 1-3, page two stanza 3-6, page 3 stanza 7; for prose versions: page 1 presented the first two text blocks, consisting of stanza 1-4 and page 2 presented the second two text blocks consisting of stanza 5-7).

Although the prose version caused one critical region to coincide with the position of the last word on the screen, which is commonly known to be a problematic area regarding eye-movement behavior, we decided to keep this structure to examine effects caused by a disruption of expected rhythm at the end of the rhythmic system (auditive gestalt) of the prose version, as well as the poem version. This decision was also based on results reported by Wassiliwizky et al. (2017), who measured skin conductance to investigate emotion and aesthetic appreciation while listening to poems and found that chills occurred at the end of line, end of stanza and end of a poem.

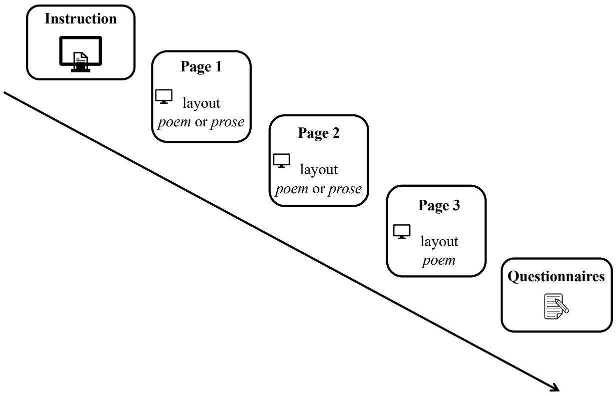

Figure 2. Illustration of experimental setup.

The three types of experimental manipulations (*meter rhyme*, *rhyme&meter*; see appendix for all stimuli) are shown *figure 3*.

The first stanza of a poem introduced its rhythm, so participants had the chance to pick it up while reading silently and to potentially build rhythmic expectations. The rhythm of each poem was closely aligned to its main metrical grid to make sure MRRL was strongly metrical (compare figure 2 in Ravignani & Madison, 2017), thus allowing for ‘quasi-isochrony’. We also added combined rhyme and metric anomalies (manipulation 4). These anomalies presumably impede the accommodation of the rhyme scheme into an ABAC pattern.

Stanzas 3 and 5 were in accordance with the rhythmic constraints so that readers might pick up the rhythm again. Manipulations (2) and (4) in stanza 7 allowed for complete deviation from the ABAB rhyme-scheme. In the present study, ABAB scheme implies perfect as well as imperfect, but acoustically close rhymes. Findings by (Knoop et al., 2019, 10f) suggest “that imperfect rhymes benefit from metered verse context” and “are harder to distinguish from perfect rhymes as distances increase”, presumably depending on the “degree of phonological similarity”.

Note that we introduced the different rhythmic deviations on the basis of the constraints named above (for details see Appendix).

PROSE LAYOUT

Wir hatten keine Kerzen bei und auch die
Taschenlampen nicht, es war ja auch ganz
einerlei, wir liefen gut auf freie Sicht. Die
Taschen waren voll bepackt mit allem was die
Welt begehrt, wir gingen Gleichschritt, fast im
Takt, doch schnell war's meiste aufgezehrt $_1$ | **aufgebraucht** $_2$

Der Hunger dennoch war vorbei, was andres war
viel lockender… Bald Nacht vom Tage ganz entzwei
und Füße trabten stockender. Die Bäume jäh schon
trennten sich und vor uns liegend Wassers Gang.
Das Ufer drüben nur ein Strich, er ging am untern
Himmel lang $_1$ | **entlang** $_3$

Da stoppten wir mit Atmung still und blickten
in die Weite hin und hörten kurz noch Grillen
schrill, doch zügig war'n sie aus dem Sinn. Denn
lichternd war's am dunklen See und wieder waren
alle da. So manches knipste schnell noch Klee
hell blinkend an, wir machten Ah! $_1$ | **Ohoh**! $_4$

Die Würmchen flogen froh im Schwarm und unsre
Augen hintendrein, beim bloßen Schauen wurd es
warm. $_1$ | **heiß.** $_2$ So darf im Mai das Glühen sein. $_1$ | **werden**. $_4$

POEM LAYOUT

Wir hatten keine Kerzen bei
und auch die Taschenlampen nicht
es war ja auch ganz einerlei
wir liefen gut auf freie Sicht

Die Taschen waren voll bepackt
mit allem was die Welt begehrt
wir gingen Gleichschritt, fast im Takt
doch schnell war's meiste aufgezehrt $_1$ | **aufgebraucht** $_2$

Der Hunger dennoch war vorbei
was andres war viel lockender…
Bald Nacht vom Tage ganz entzwei
und Füße trabten stockender

Die Bäume jäh schon trennten sich
und vor uns liegend Wassers Gang
Das Ufer drüben nur ein Strich
er ging am untern Himmel lang $_1$ | **entlang** $_3$

Da stoppten wir mit Atmung still
und blickten in die Weite hin
Und hörten kurz noch Grillen schrill
doch zügig war'n sie aus dem Sinn

Denn lichternd war's am dunklen See
und wieder waren alle da
So manches knipste schnell noch Klee
hell blinkend an, wir machten Ah! $_1$ | **Ohoh**! $_4$

Die Würmchen flogen froh im Schwarm
und unsre Augen hintendrein
beim bloßen Schauen wurd es warm $_1$ | **heiß** $_2$
So darf im Mai das Glühen sein $_1$ | **werden** $_4$

Figure 3. Illustration of poem layout and prose layout: (1) original text, (2) *rhyme anomaly*: substitution of rhyme with original number of syllables, (3) *metric anomaly*: change of prominent metrical figure by adding one or two syllables with rhyme being maintained, (4) *rhyme and metric (rm) anomaly*: change of prominent metrical figure by adding one or two syllables, with substitution of rhyme.

Rhyme anomaly. Since the first stanza introduces the ABAB scheme, readers may use it as a default for the upcoming stanzas. Rhyme anomalies such as *begehrt*/*aufgebraucht* instead of *begehrt*/*aufgezehrt* do not violate a potentially superimposed regular beat distribution but they may collide with the expected rhyme scheme. This holds true for imperfect rhymes, too.

Metric anomalies were construed by adding one to two syllables, disturbing the grouping structure of the previous syllabic material in the stanza. This was done by e.g., violation of expected stress/accent, by missing and/or delayed accent or by preponed and/or added accent. Examples are e.g., *Gang*/*lang* vs. *Gang*/*entlang*, leading to an additional floating stress moment, or *grad*/*Waldesnaht* vs. *grad*/*Waldesziernaht*, leading to preponed stressing of "zier" and stress diffraction on the last syllable "naht". Adding a syllable could also shift the projected number of beats if introducing one more syllable which requires stress, thus locally disturbing the overall stress distribution within the stanza.

Metric & rhyme anomaly should most clearly lead to irritation within the overall rhythmically structured 'gestalt', either by realizing possibilities listed above combined with deviation from the rhyme scheme, or, by implying a stress clash, e.g. gegeben/(gut) durchleben vs. gegeben/(gut) überstehen (see Appendix for further details). However, our focus was not to analyze the different sub-types of metrical anomalies or rhyme anomalies, but more so the general eye-movement reactions elicit by the anomalies.

The corresponding prose version which includes experimental manipulations had the same pattern (adjusted interpunction marked red). Content-wise, both prose versions, original and manipulated, were in line with the corresponding poem versions. Changes in prose versions were undertaken for two purposes: a) line breaks should not coincide with the position of pre-rhymes, and b), when line breaks coincided with clause boundaries in the poem layout, interpunction and capitalization was adjusted to preserve the clause structure.

Seven poems were composed by the first author specifically for the purpose of the experiment. One more poem was an original, "Auf hohem Gerüste" (Ringelnatz, 1997, p. 63; excluded from data analysis). Hence, the stimuli have not been used in previous research. These seven poems followed a preset poetic rhythm structure as close as possible. This was obtained by adherence to the rhythmical matrix of classical originals, i.e., 1) *Dancing Queen*, 2) *Flüstern*, as in "Der Pilgrim" by Friedrich Schiller, 3) *Klimawandel* as in "Der Wanderer in der Sägemühle" by Justinus Kerner, 4) *9 Leben* as in "Auf hohem Gerüste" by Joachim Ringelnatz, 5) *Normal* as in "Am Waldessaume träumt die Föhre" by Theodor Fontane, 6) *Im Hüteland* and 7) *Glühwürmchen* were authored following preponderantly the rhythmic matrix (rhyme, meter, phonological relatedness) of those named above. They all had to rhyme according to the ABAB-scheme, which could also include imperfect, yet acoustically close rhymes.

The semantic field of words was chosen from commonly known topics such as nature, summer, youth, desperation, etc. Poems mostly contained familiar and high frequency words, such as *luck, stars, sky, forest, breathing*, etc., function words as well as some low frequency or antiquated words, and neologisms.

The seven new poems included parallistic dictions and a higher level of difficulty (Castiglione, 2019; Yaron, 2002, 2008) compared to "Auf hohem Gerüste" by Ringelnatz, i.e., they presented a moderate number of stylistic devices such as assonances, alliterations, comparisons, e. g. "die Sterne wie Glitzerstuck am Himmel" (the stars like glittering stucco in the sky) or neologisms, e.g. "Hügelzwerg" (hill dwarf), etc. We did not exclude any non-standard syntactic patterns, because word order is an important stylistic feature contributing to the multilayered meaning and rhythm construct of a MRRL-poem (Schrott & Jacobs, 2011).

Although stimuli were written in a sound-familiar metrically regular and rhymed style (such as quatrains, nursery rhymes, etc.), the choice of words and the occasionally complicated syntax should prohibit complete and deep sentence comprehension. At the same time, we expected readers to grasp the narrative of a poem quickly (Castiglione, 2017), i.e., global comprehension of content. For this reason, we assumed fluent reading, which in turn was presumed to enhance rhythmic subvocalization. Also, participants were not allowed to move back to earlier pages, which also made full sentence comprehension within the course of a poem more difficult. This ensured that rhythm became a more salient feature.

Procedure

The experiment was designed and conducted in the Cognitive Science eye-tracking laboratory at the University of Freiburg. Participants were asked to sit at a desk across from the eye-tracker table, to read a brief information sheet and to give written informed consent prior to the experiment. Next, they were asked to sit in front of the eye-tracker. Body position adjustment and camera setup (calibration and validation) were undertaken.

The recording session started with a short instructional text on the screen (see appendix for exact wording). Its purpose was to acquaint participants with reading in front of an eye-tracker with their head in a head-and-chin rest. The text informed participants about the fixation cross and the space bar so that they would know how to proceed to the following page. It also invited them to be curious about the content and asked for their attention to the upcoming texts. No instruction for reading speed was given. Stimuli were presented in randomized order for each participant. At the beginning of each trial participants had to fixate on a cross at the position where the first word of the item would appear and press the space key. Once they did so, the first page appeared on the screen. When they finished reading a page, participants could move to the next page by again pressing the spacebar. No option for moving back to the previous page was provided. After they had finished reading the last page of a trial, the next trial was indicated by the next fixation cross.

After the recording session was finished, participants were asked to sit again at the first desk and to fill out questionnaires (1. processing of stimuli, 2. reading habits, 3. *A short Questionnaire to Assess Musical Activity*, MusA (Fernholz et al., 2018)). Participants were allowed to ask questions with regard to proper understanding of questions (such as: “Does this question apply to all texts?”). Answers were only given when necessary. Otherwise, participants were invited to read again closely and to give an answer that would seem appropriate to them.

After finishing all three questionnaires, a short feedback interview took place with questions like “Did you notice anything special about the texts?”. If key words like *rhythm*, *expectation*, *(inner) voice* or semantically or thematically close words were part of an answer, participants were asked to specify what they meant by these words. In addition, they were asked whether they had noticed something about their eyes or whether they had inwardly heard something like a voice during reading silently or not. If the latter was confirmed, they were asked to try to explain a possible function of that inner (reading) voice. Notes of answers were jotted down. Questionnaire and interview data has not been included in the analysis and will be discussed elsewhere.

Data Analysis

Fixation reports of the raw data were generated using the SR-Research Data-Viewer. Blink durations were not included in fixation durations. Fixations occurring directly before or after a blink were not excluded from the data set. Rectangular interest areas (IA) were defined automatically around each word on a page. Every computational step from here, including interest area assignment, was taken in the R programming language. The code is available upon email request.

For each fixation, we assigned an IA based on the fixation’s x and y coordinates. Fixations’ start times were used to identify the page - one out of three - that was read. The completed fixation reports were then transformed into IA-reports, with each row representing a consecutive IA/word in an item, including variables for eye tracking measures, lexical features and other IA related variables that would potentially affect reading measures, including the design factors.

Word reading time measures, especially in longer texts, are affected by many variables that are not in the main focus of our study. However, to control for these variables, we consider it mandatory to account for their influence. This should be done on as many data as possible, namely on all words in the texts, with the exception of the first word.

For the data analysis of the critical IAs, we hence chose a two-stage approach, where the analysis of critical IAs was based on residuals derived from all IAs.

However, we were also interested in general eye-tracking signatures of subvocalization on areas other than the critical IAs. We therefore chose to analyze the IA-reports in two parts (except for skipping probability and load-contributions; see *figure 4*).

Part 1 focused on the reading of the critical IAs themselves. This analysis has been carried out in two stages. In stage 1 we fitted a base model over all IAs (words). The purpose of the base model is to eliminate all effects that are not (related to) the design factors, which are included in the main model, namely *layout, anomaly_type* and *MRRL_version*. The base model includes a wide variety of general predictors that are known or very likely to influence eye-movements and word reading times. Among those were *i.* lexical features, such as word length, frequency (Just & Carpenter, 1980; Kliegl et al., 2004; Schuster et al., 2016), and the word category (noun, verb, adjective, closed class words), *ii.* structural features, such as whether an interest area (word) occurred at the end or the beginning of a line (Koops van ’t Jagt et al., 2014) or verse (rhyme indicator) (Carminati et al., 2006). Finally, we also included *iii.* oculomotor behavior variables, such as whether or not a first pass regression is launched, and gaze durations of the predecessor word. These variables can strongly affect all duration measures independently of our design factor manipulations and should thus be accounted for, either in the base or main model. Accounting for them in the base model has the advantage of almost completely detaching them from the critical IAs, where the effect of the design variables should be as pure as possible.

The base model was only fitted to produce residuals (Trueswell et al., 1994), which were then used as the response variables in the second stage models. Using residual reading times is a common technique to account for, and eliminate, irrelevant influences before looking at the effects of the design factors. Note that the base-model was fitted across *all* interest areas.

The residuals were then used in stage 2 (i.e., the main model) to analyze a reduced data set, where all but the critical interest areas (IAs) were excluded. Because distractor influences were eliminated in stage 1, the main model only included the design factors as fixed effects predictors. Critical interest areas were those target words that have been manipulated in the experimental conditions, i.e., replaced with other words inducing a meter or rhyme anomaly, or both.

		PART 1: critical IAs	PART 2: other IAs
S T A G E	1	**Base model** *Interest areas* all IAs *Predictors* generic variables influencing word reading time, incl. lexical variables	**Complete model** *Interest areas* all IAs except critical *Predictors* *base model*, plus design variables MRRL_version & layout
	2	**Main model** *Interest areas* critical IAs only, *Predictors* design variables, *Response variable* base model residuals	

Figure 4. Scheme of analysis.

The two-stage approach was chosen for two reasons: First, we could include a plethora of variables influencing reading times in the base model without sacrificing power in the main model. The main model could thus be based on residual eye-tracking parameter values that were fitted over the entirety of the poems, consisting of about 160 words each. Had we chosen to include all predictors in a single model, not only would we have lost power by analyzing only five interest areas (words) per poem. Secondly, estimates of lexical variables would have been obscured by any manipulation that disrupts reading, particularly so the anomalies. Only results from the main model of part 1 will be reported.

Complete model. However, we were also interested in how our design manipulations affected reading in general - not only at the target words, but throughout the entire poem. Hence, part 2 focused on the effects of our manipulations on all but the critical IAs. This *complete model* included all predictors from the base model in stage 1, plus the design factors *layout* (layout: poem vs. prose) and *MRRL_version* (consistent vs. inconsistent). Factor *anomaly_type* was not included, because it was only defined for critical IAs, as all manipulated stimuli contained all three types of anomalies (anomaly_type metric, anomaly_type rhyme, anomaly_type r&m).

For all analyses, *linear or logistic mixed effect regression* models were fitted using the *lme4*-package (Baayen, 2008; Bates et al., 2015) in R (R Core Team, 2020). Further packages used were *LMERConvenience Funcions* (Tremblay, A. & Ransijn, J., 2015), *lmerTest* (Kuznetsova et al., 2017), and *multcomp* (Hothorn et al., 2008).

For the complete models, we used stepwise elimination to yield a minimal model, which only included predictors that significantly increase the model quality. For this, we used the function step() from the *lmerTest* package, which applies backward elimination of random-effect terms followed by backward elimination of fixed-effect terms in linear mixed models.

The variance inflation factors of all predictors in both the main and complete models were below 5.

The main model of part 1 included the study design factors *layout* (*layout* with levels *poem* vs. *prose*), *MRRL_version* (with levels *inconsistent* vs. *consistent*), and *anomaly_type* (levels *metric* vs. *rhyme* vs. *r&m* for *rhyme+metric,* respectively) and all interactions between the three factors.

In both the base and the main model, intercepts for participants and items were included as random factors. The rationale for this is the different sets of IAs and predictors in both models. Some readers might react to anomalies and layout manipulations differently, resulting in estimate variance, even after general reading measures have had normalized across all IAs. Also, stimulus manipulations might have different effects in different items. Furthermore, slopes for *word length* and *frequency* were added in the base model, and the slope for *MRRL_version* in the main model.

Variables in the base and complete model. We included three types of variables in both the base and the complete model: lexical, structural, and oculomotor variables.

Lexical variables. We computed five *lexical features*: 1. word category annotated *cat* (labeled *catC*, *catA*, *catN*, *catV*; which identified levels *closed class,* *adverb/adjective, noun*, *verb*), 2. *word length*, i.e., the number of characters for each word (*word_length)* and 3. *log word frequency* (*log.freq*) based on the DeReWo-2014 corpus-based word lists (Belica, C., Kupietz, M., Lüngen, H., & Perkuhn, R., 2012).

We computed 4. the *consonant vowel quotient* (*cvq*), as an indicator of pronounceability (Kraxenberger et al., 2018; Lee et al., 2002; Rayner & Pollatsek, 1989; Xue et al., 2019). The calculation was based on letters rather than sounds. For German, a high level of consonants is assumed to impede pronunciation, as can be experienced in tongue twisters (e.g. "Schlickkrebskriechgang" / "Schlickkriechkrebs-schleichgang"). We also added the consonant vowel quotient of the succeeding word (cvq.p1) as an indicator of parafoveal processing of phonological/pronunciation information.

Finally, 5. the *number of syllables* (*syllables*) of a word were computed as an estimate of how long it would take to be spoken. Naturally, number of syllables and the number of characters (*word_length*) of words are highly correlated (.84, see table 1). We therefore computed residualized number syllables (*res.syllables*) in a simple regression over word-types, where syllables were predicted from word length. *Res.syllables* is thus independent of word length and reflects pronunciation more purely. In earlier research, syllable number has been shown to influence skipping, but no effect on reading time measures beyond word length was found in normal reading (Fitzsimmons & Drieghe, 2011). Hence, we would consider any such effect in our results a strong indicator for an eye-voice-span synchronization induced by MRRL-language.

Table 1. Correlation matrix of lexical variables

Variables	1	2	3	4
1. word_length	—			
2. syllables	.84	—		
3. log.freq	-.58	-.51	—	
4. cvq	.15	-.26	-.06	—

Also, since the *cvq* turned out to be highly negatively correlated with *res.syllables*, we computed the residual cvq (*res.cvq*) by predicting the *cvq* from both *res.syllables* and *word_length* in a linear regression model over word types.

Structural variables. In addition, we computed variables related to particular IA-positions that are known to influence reading, such as the beginning (*BOL*) or end of a line (*EOL*).

Furthermore, we included the variables beginning of verse (*BOV*) and end of verse (*EOV*). Although *EOV*s coincide with *EOL*s in poem layout, they do not necessarily do so in prose layout. The ending of a verse signals an end point of an important (rhythmic) unit and could thus influence subvocalization, e.g. by triggering a pause, independent of a visual line break.

We also included *page* number, the running word number on a single page (*wpos*), and the interaction between the two in order to capture adaptation effects throughout reading a complete item. To account for potential practice or fatigue effects we included the variable *trial* (values 1 to 8), encoding the presentation order of trials throughout the experiment, i.e., the position number of each trial in the experiment.

Oculomotor variables. To account for potential preview and spill-over effects we included the gaze durations of the predecessor word (*gaze_pre.word*) as a linear predictor. Because first pass duration measures can vary considerably depending on whether first pass reading is followed by a regressive saccade, we also added the binary predictor *first_pass_regression*.

Eye tracking parameters. Before we computed eye-tracking measures from the fixation reports, all single fixations on an IA shorter than 40 milliseconds were treated as overshoots and assigned to the previously fixated IA. Data cleaning, including outlier elimination, was done completely automatically. For each IA, we computed *first fixation durations* (*FFD*), *single fixation durations* (*SFD*; equaling *FFDs*, but excluding all cases with more than one fixation during first pass), *gaze duration* (*GAZE*, the sum of all fixations on the target IA during first pass), *regression path duration* (*RPD*), the sum of all fixation durations during first pass plus - if the first pass is followed by a regressive saccade - all fixation durations on predecessor IAs, until a saccade goes past the target IA (Konieczny et al., 1997), right bounded reading time (*RBRT*, the sum of all fixation durations on the target IA until a saccade goes past the IA), *total reading times* (*TRT*, the sum of all fixations on an IA), and *second pass reading time* (*SPRT*, computed as *TRT* minus *GAZE*). All first pass measures (*SFD*, *FFD*, *RD*, and *RBRT*) required the first fixation resulting from a progressive saccade. Also, we analyzed conditionalized times, meaning that zero values were treated as missing values. For data analysis, all time-based parameters were logarithmized.

In addition to these reading time measures, we computed variables coding whether or not a word has been skipped (*SKIP*).

Before model fitting, we calculated overlaps and correlations between the eye tracking parameters (see table 2). Because single fixation durations (*SFD*) are a subset of first fixation durations and first pass reading times, their correlation must equal 1. All other measures – with the exception of *SPRT* (second pass reading times) and both *SFD* and *GAZE* – are significantly correlated with each other (p<.001), albeit to a varying degree.

Table 2. Correlations between common eye-movement parameters.

Variables	1	2	3	4	5	6	7
1. SFD	—						
2. FFD	1	—					
3. GAZE	1	.47	—				
4. RPD	.06	.08	.21	—			
5. RBRT	.41	.33	.74	.51	—		
6. TRT	.28	.25	.57	.42	.78	—	
7. SPRT	.01	.01	.03	.28	.49	.87	—

Single and first fixations are a subset of fixations that constitute gaze durations, and therefore their correlation is 1. However, since single fixation durations (*SFD*) and *GAZE* share only 74.7% of the data points, we will report results from both model fits. First fixation durations (*FFD*), on the other hand, will be ignored. Right bound reading times (*RBRTs*) are highly correlated with regression path durations (*RPD*), so they will be ignored, too. We also ignored second pass reading times (*SPRTs*), because they are highly correlated with total reading times (*TRT*, .87). The remaining measures should suffice to tap into early and later processing stages.

Total reading times are a combined measure of first pass and later processing. Therefore, there will be an overlap with *GAZE* and single fixation durations (*SFD*), but any deviations would suggest later stage processes. Total reading times (*TRT*) are thus considered a measure of overall processing difficulty.

Finally, we computed *Load Contributions* (Konieczny et al., 2000) as a measure of selective re-reading. *Load contributions* (*LC*) measures the time spent re-reading (sum of all fixations on) a previous region in the regression path of a later region. This measure is of particular relevance, because we are interested in whether the eyes re-fixate the pre-rhymes in cases of meter and rhyme anomalies.

Before each stage, and for each response duration variable, extreme values were eliminated. We first identified extreme values by using the function boxplot() with range 3. Hence, outliers were defined as values beyond the most extreme data point which is no more than three times the inter-quartile range from the box.

Then we fit the *base model* (stage 1), and again - in the same way - identified and eliminated extremes in the residuals. The *base model* was fit a second time and the resulting residuals were finally merged back into the dataset. From here on, only the critical interest areas were used to fit the *main model*.

For duration variables, we fit linear mixed effects regression models, using the function *lmer()* from the *lme4 R*-package (version 1.1-21; Bates et al., 2015, p. 4). The binary variable skip was analyzed with *logistic* mixed effects regression, using the function *glmer()*.

For all model fits, we used *sum contrast coding*, creating predictors for all but the last level of any categorial variable and assigning 1 to the corresponding level for each comparison as well as -1 to the last level for all comparisons. Remaining levels were coded 0 (*table* 3).

In *sum coding*, the intercept represents the grand mean, and each contrast represents a comparison of a factor level mean to the grand mean. Therefore, all effects are independent of each other. Hence, simple contrasts can be interpreted similar to main effects in ANOVAs – even when the predictor also occurs in interaction terms in the model.

Table 3. Sum contrast coding for variable anomaly_type.

	metric	rhyme
Metric	1	0
Rhyme	0	1
r&m	-1	-1

P-values for linear mixed models were estimated with Satterthwaite's approximation of degrees of freedom, using the *lmerTest R*-package (version 3.1-0, (Kuznetsova et al., 2017).

We will first present reading time results, starting with the *main model* and continuing with the *complete* *model*. *Skipping* results will be presented next, and lastly, we will present *Load Contribution* results.

Extreme values. For *single fixation durations*, seven extreme values were filtered from raw data (0.02%). Additional 32 data points (0.08%) were eliminated as outliers from the complete model, and 35 (0.09%) from the base model. No extreme values were excluded after fitting the main model.

For *gaze durations* (first pass reading times), seven extreme values were filtered from raw data (0.02%). No extreme values were excluded from the data set for the base, the complete and the main model.

For *regression path durations*, 32 extreme values were filtered from raw data (0.08%), 10 data points (0.03%) were eliminated as outliers from the complete model, 13 data points (0.03%) from the base model. No extreme values were excluded from the dataset for the main model.

For *total reading times*, 2 extreme values were filtered from raw data (0%). One additional data point (0%) was eliminated as outlier from the complete model, and 2 data points (0%) were as outliers eliminated for the base model. No extreme values were excluded from the dataset for the main model.

Results

Main Model. We predicted that metrical anomalies induce disruptions, if the metrical structure of the poems was recognized and the anomaly was hence experienced as diverging. We also predicted that the poem layout facilitates the capturing of the rhythmic gestalt, enhancing potential effects of metrical anomalies.

The results of the linear mixed effects model fits of the *main model* (table 4) show a fairly robust pattern across eye-tracking measures (see *figures 5 to 8*). In poem layout, metric anomalies resulted in increased fixation and reading times compared to the metrically consistent version. This amounts to a significant three-way interaction of factors *layout*, *MRRL_version*, and *anomaly-type metric* for measures *GAZE*, *RPD*, and *TRT (SFDs* were only marginally reliable: p=.066 - two way test), on top of the two way interaction of *layout* and *anomaly_type metric* for measures *SFD*, *GAZE* and *RPD*, and a main effect of *MRRL_version* for measures *RPD* and *TRT*.

Post-hoc contrasts between *inconsistent* and *consistent MRRL_versions* of *metrical* anomalies in the *poem* layout turned out to be significant for *SFDs* (directional hypothesis), *GAZE, RPD*s, and *TRTs* (see table A). Also, in the *poem layout* (table C), post hoc contrasts between inconsistent MRRL_versions of *metric* and *rhyme* anomalies (*SFD*, *GAZE*, *RPD*), as well as *metric* and *r&m* anomalies turned out to be significant (all measures). *Metric anomalies* (*MRRL_version inconsistent*) also elicited longer reading time in *poem* than in *prose layout* (see table B, all measures).

We also expected effects of rhyme violations to be stronger in poem layouts, if the layout was necessary to recognize the poetic form. On the other hand, identifying the pre-rhymes may be more demanding in prose layout, where visual cues to their positions are lacking.

In *prose layout*, *rhyme* anomalies notably triggered longer reading times, resulting in reliable three-way interactions of *layout*, *MRRL_version*, and *anomaly_type rhyme* in *SFDs*, *GAZE*, and *RPDs*, on top of a two-way interaction between *layout* and *anomaly_type rhyme* for *SFDs*, and the aforementioned main effect of *consistency* (*MRRL_version*). Simple post-hoc contrasts between the *inconsistent* and *consistent* version in that condition were significant (directional hypothesis) for gaze durations and *RPDs* (see table A). Combined *metric* and *rhyme* (*r&m*) anomalies (*MRRL_version inconsistent*) in *prose layout* elicited smaller *GAZE* durations but larger *RPDs* than their *consistent* counterparts. This pattern suggests that *r&m* anomalies in prose layout triggered early regressive saccades during first pass reading, resulting in shorter *GAZE* durations and longer *RPDs*.

Table 4. Linear mixed effects regression coefficients of the *main model*. Time measures are *residualized and logarithmized.*

	eye.param	*Estimate*	*se*	*df*	*t*	*p*
(Intercept)	SFD	0.07	0.033	8.8	2.132	=.063
	GAZE	0.093	0.044	10	2.085	=.064
	RPD	0.144	0.04	20.7	3.6	=.002
	TRT	0.114	0.045	11.8	2.567	=.025
anomaly_type metric	SFD	0.059	0.028	410	2.104	=.036
	GAZE	0.063	0.029	546.9	2.172	=.03
	RPD	0.037	0.03	37.7	1.213	
	TRT	0.041	0.029	122.3	1.402	
anomaly_type rhyme	SFD	-0.018	0.026	48.6	-0.692	
	GAZE	-0.004	0.033	44.7	-0.12	
	RPD	0.015	0.034	37.2	0.447	
	TRT	0.028	0.031	38.8	0.901	
MRRL_version inconsistent	SFD	0.014	0.02	48	0.696	
	GAZE	0.014	0.019	86.9	0.703	
	RPD	0.046	0.021	49.9	2.239	=.03
	TRT	0.056	0.018	217.8	3.03	=.003
layout poem	SFD	0.026	0.017	219.4	1.473	
	GAZE	0.028	0.019	129.4	1.479	
	RPD	0	0.019	99.2	0.008	
	TRT	-0.003	0.019	46.8	-0.158	
anomaly_type metric:MRRL_version inconsistent	SFD	0.025	0.028	606.5	0.885	
	GAZE	0.045	0.029	1056.1	1.556	
	RPD	-0.006	0.028	1025.5	-0.217	
	TRT	0.033	0.028	1065.6	1.165	
anomaly_type rhyme:MRRL_version inconsistent	SFD	0.004	0.022	633.8	0.172	
	GAZE	0.023	0.024	1066.3	0.939	
	RPD	-0.012	0.024	1027.7	-0.501	
	TRT	-0.01	0.023	1091.7	-0.429	
anomaly_type metric:layout poem	SFD	0.077	0.028	605.5	2.751	=.006
	GAZE	0.082	0.029	1059.1	2.861	=.004
	RPD	0.1	0.028	1041.1	3.531	<.001
	TRT	0.025	0.028	1070.2	0.911	
anomaly_type rhyme:layout poem	SFD	-0.051	0.022	629.4	-2.256	=.024
	GAZE	-0.029	0.024	1066.7	-1.174	
	RPD	-0.024	0.024	1025.8	-1.018	
	TRT	-0.006	0.023	1090	-0.239	
MRRL_version inconsistent:layout poem	SFD	0.018	0.017	621.2	1.018	
	GAZE	0.022	0.018	1081.3	1.208	
	RPD	0.013	0.018	1061.4	0.742	
	TRT	0.022	0.018	1116.7	1.206	
anomaly_type metric:MRRL_version inconsistent:layout poem	SFD	0.051	0.028	595.4	1.842	=.066
	GAZE	0.059	0.029	1059.4	2.044	=.041
	RPD	0.065	0.028	1036.9	2.305	=.021
	TRT	0.057	0.028	1068.4	2.039	=.042
anomaly_type rhyme:MRRL_version inconsistent:layout poem	SFD	-0.047	0.022	626.8	-2.107	=.036
	GAZE	-0.062	0.024	1067.8	-2.556	=.011
	RPD	-0.05	0.024	1024	-2.123	=.034
	TRT	-0.036	0.023	1089.7	-1.52	

Note. For SFD, the number of observation was 699, the conditional R^2 was 0.118 and the marginal R^2 was 0.0213. For GAZE, the number of observation was 1143, the conditional R^2 was 0.129 and the marginal R^2 was 0.0246. For RPD, the number of observation was 1125, the conditional R2 was 0.176 and the marginal R2 was 0.0265. For TRT, the number of observation was 1188, the conditional R2 was 0.127 and the marginal R2 was 0.0188.

*Model specification: res.LOG.<duration> ~ anomaly_type * MRRL_version * layout + (1 | vp) + (0 + MRRL_version + layout + anomaly_type | vp) + (1 | item)*

Post-hoc contrasts between *layout poem* and *prose* for combined *metric* and *rhyme* (*r&m*) anomalies (see table B) show shorter *RPDs* for layout poem for both versions (*consistent*/*inconsistent*) and overall longer ones for *layout prose*. One possible explanation is that this result is an artefact which can be traced back to the positions of the *r&m* anomalies. First, because of the presentational conditions (see *figure 3* or *Appendix*), in *prose layout, r&m* anomalies appeared on page 2. As a result, here, readers were confronted with more text material, which could have elicited longer *RPDs* for both, *MRRL_version inconsistent* and *consistent*. Second, because in poem stimuli presentation readers were not able to jump back from page 3 to page 2, in the *poem layout*, less text material could be re-read. This is applicable for both versions (*consistent*/*inconsistent*) and most likely accounting for the shorter *RPDs* compared to prose layout.

Table 5. Main model post-hoc contrasts. Time measures are residualized and logarithmized

MRRL_version (A)	*contrast*	*eye.param*	*Estimate*	*SE*	*df*	*t.ratio*	*p*
layout = poem, anomaly_type = metric	inconsistent - consistent	SFD	0.2153	0.1194	362	1.804	=.072
		GAZE	0.27800	0.1108	813	2.509	=.012
		RPD	0.23684	0.1094	746	2.164	=.031
		TRT	0.3350	0.1054	867	3.179	=.002
layout = prose, anomaly_type = rhyme		GAZE	0.15293	0.0831	514	1.840	=.066
		RPD	0.14291	0.0826	433	1.729	=.085
layout = prose, anomaly_type = r&m		GAZE	-0.15909	0.0780	458	-2.039	=.042
		RPD	0.13070	0.0783	387	1.670	=.096
layout (B)	*contrast*	*eye.param*	*estimate*	*SE*	*df*	*t.ratio*	*p*
anomaly_type = metric, MRRL_version = inconsistent	poem - prose	SFD	0.34350	0.1218	406	2.821	=.005
		GAZE	0.381739	0.1058	818	3.608	<.001
		RPD	0.3561	0.1045	807	3.408	<.001
		TRT	0.2485	0.1036	811	2.400	=.017
anomaly_type = r&m, MRRL_version = inconsistent		RPD	-0.1525	0.0758	461	-2.011	=.045
anomaly_type = r&m, MRRL_version = consistent		RPD	-0.1484	0.0788	501	-1.884	=.060
anomaly_type (C)	*contrast*		*Estimate*	*SE*	*df*	*t.ratio*	*p*
MRRL_version = inconsistent, layout = poem	metric - rhyme	SFD	0.323769	0.1056	248	3.065	=.007
		GAZE	0.3197	0.0968	288	3.303	=.003
		RPD	0.26612	0.0964	255	2.762	=.017
	metric - r&m	SFD	0.312746	0.1040	244	3.006	=.008
		GAZE	0.4242	0.0944	343	4.496	<.001
		RPD	0.31847	0.0943	303	3.379	=.002
		TRT	0.28156	0.0906	373	3.106	=.006

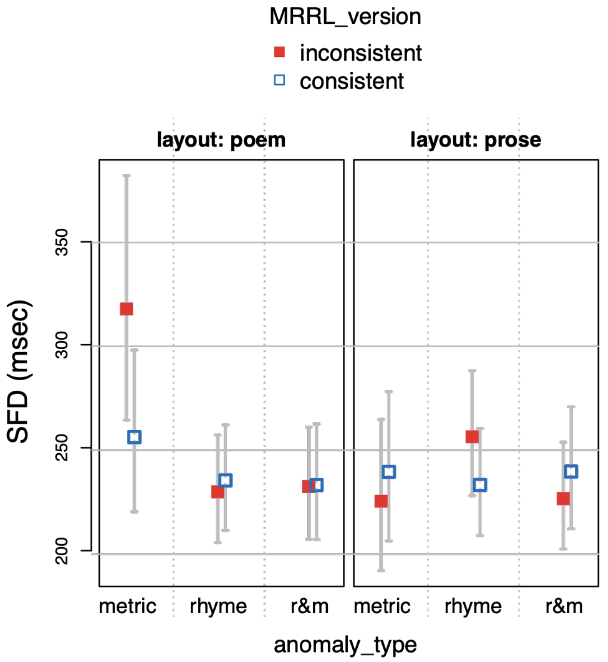

Figure 5. Single fixation durations (SFD) as a function of layout, MRRL_version and anomaly_type.

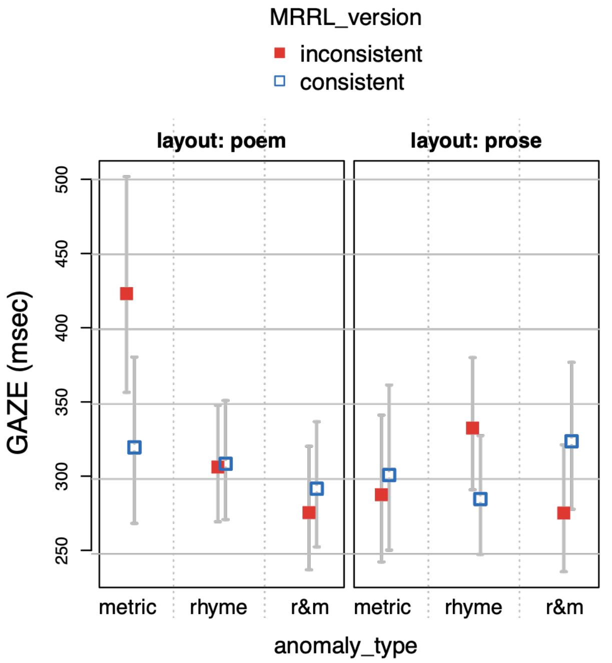

Figure 6. Gaze durations (GAZE) as a function of layout, MRRL_version and anomaly_type.

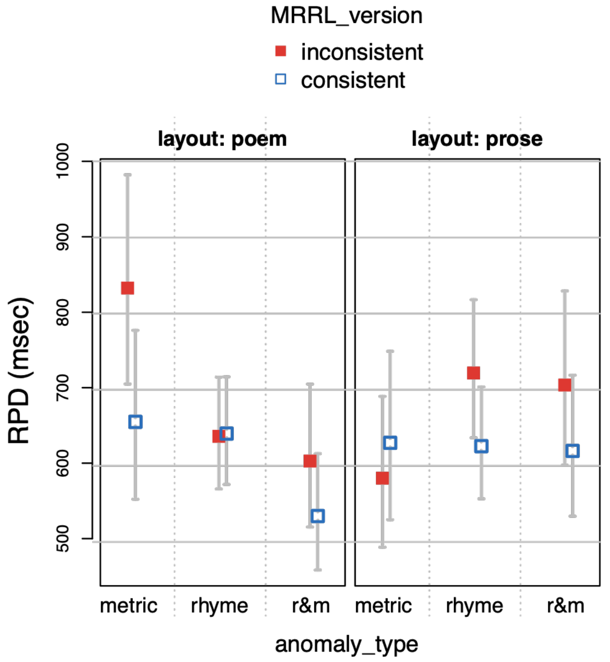

Figure 7. Regression path duration (RPD) as a function of layout, MRRL_version and anomaly_type.

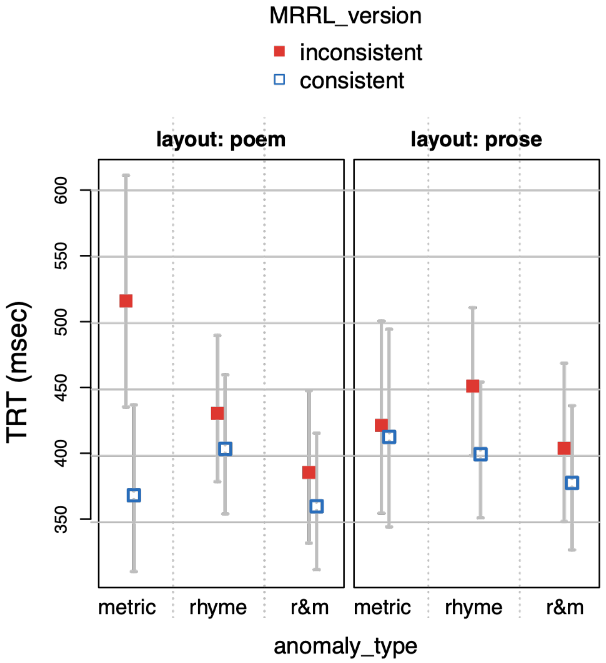

Figure 8. Total reading time (TRT) as a function of layout, MRRL_version and anomaly_type.

The whiskers for figure 5-8 indicate 95% confidence intervals. Raw reading times were back-transformed from residual logarithmized model estimates by first adding the base model intercept and then applying the exp-function.

Discussion

For the *main model*, all four eye-tracking measures (see *figures 5 to 8*) revealed a high sensitivity to the *metric anomaly* in poem layouts, resulting in increased fixation and reading times. This finding suggests that the poem layout is mandatory to detect the metric anomaly.

Thus, when participants read poems, they pick up the overall prominent (linguistic) metrical grid, and their expectations get disrupted when the metric scheme is broken. Increased reading times on the word that introduces the anomaly in poem layouts thus indicate that readers apparently look for a solution on that very word. This is a clear early indicator for rhythmic subvocalization.

Interestingly, the rhyme anomaly effect for *GAZE* and *RPD*s in prose layout suggests that readers were expecting rhyme words to some extent, presumably elicited by the MRRL-structure. Here, rhyme anomalies have resulted in hesitations that signal disrupted expectations. In poem layout, however, rhyme anomalies have not disrupted reading, presumably because the strict poetic form facilitated the adoption of a different, yet common, rhyme scheme ABAC.

R&m anomalies in prose layout lead to shorter *GAZE* durations but longer *RPD*s, indicating that specifically the combined anomaly triggered early regressive saccades during first past readings. The seemingly contradicting results in this condition for *GAZE* and *RPD*s highlight the importance of interpreting eye tracking measures in relation to each other and not independently from one another.

Complete Model

Generic variables. As expected, more frequent words (*log.freq*) elicited significantly shorter reading times in all four variables (*SFD, GAZE*, *RPD,* and *TRT,* see *table 6*). *Word_length* also showed a significant increase of reading times for longer words in all variables except *SFD*s.

Among the pronunciation-related variables, we found a significant effect of residual number of syllables (*res.syllables*) in all four measures. The effect indicates a strong impact of subvocalization on reading. Even *SFD*s, which showed no reliable effect of word_length, were significantly increased for *res.syllables,* suggesting a closer link of *SFD*s to pronunciation rather than to visual word processing in our study.

Number of syllables (*res.syllables*) did also interact with MRRL-version in early measures (*SFD* and *GAZE*) indicating that words with more syllables were fixated even longer when anomalies were present in the poem.

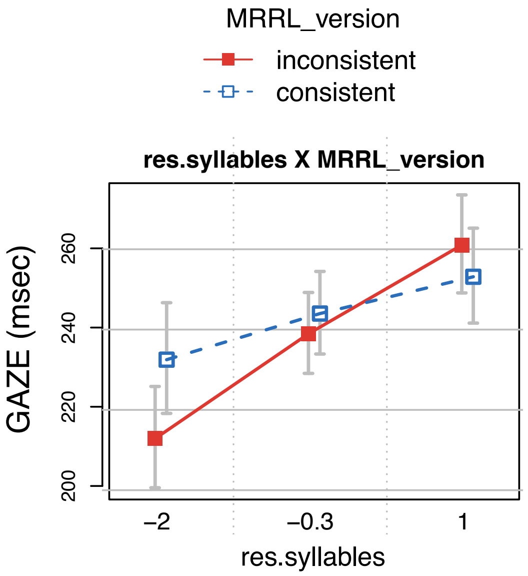

Figure 9. Gaze durations (GAZE) as a function of (residual) number of syllables and MRRL_version. The whiskers indicate 95% confidence intervals. Raw reading times were back-transformed from the logarithmized estimates.

The residual *consonant vowel quotient* (*res.cvq*), included as another indicator of subvocalization, did not turn out significant nor did any interaction with *res.cvq*.

Reliable effects of *first_pass_regression* indicate shorter *SFD*s and *GAZE* durations, whenever first pass reading ended in a regressive saccade, consequently increasing *RPD*s and *TRT*s. Line endings (*EOL*+) were read reliably faster in all four measures, whereas line beginnings (*BOL*+) were read slower than words in other line positions. Reliable effects of *trial* and *wpos* in *TRT*s indicate a speed-up for later experimental trials and throughout a single page, respectively, suggesting a practice effect or adaptation. The acceleration of *TRT*s on a single page (*wpos*) was even stronger on later pages, as indicated by a reliable interaction of *wpos* and *page.*

One might argue that the speed up towards the end of a trial contradicts the principle of isochronicity. In our view, this is not the case. One can read/recite ’empirically isochronal‘ (Ravignani 2017, 2019) along with a metrical grid and speed up or slow down reading tempo, as long as the inferred 'beat' is evenly distributed in a certain time window (similar to speeding up/slowing down in musical piece), such as a verse or a stanza.

Word *category* also showed reliable effects in all measures. Effects of these generic variables were to be expected and confirm the accuracy and soundness of our measurements.

Higher gaze durations of the previous word (*gaze_pre.word*) elicited a positive effect on single fixation durations, suggesting a short processing spill-over as well as a negative effect on the *GAZE*, *RPD*s and *TRT*s suggesting a preview effect.

While the presence of anomalies (*MRRL_version inconsistent*) did not alter reading over all, measures representing late processing, *RPD*s and *TRT*s, showed significantly slower reading in *poem layout*, and a slight acceleration in *TRT*s towards the end of the experiment (interaction *layout* by *trial*), in particular for inconsistent poem versions (interaction *MRRL_version* by *layout* by *trial*). TRTs apparently reflect the adaptation to the design manipulations very well (*figure 10*).

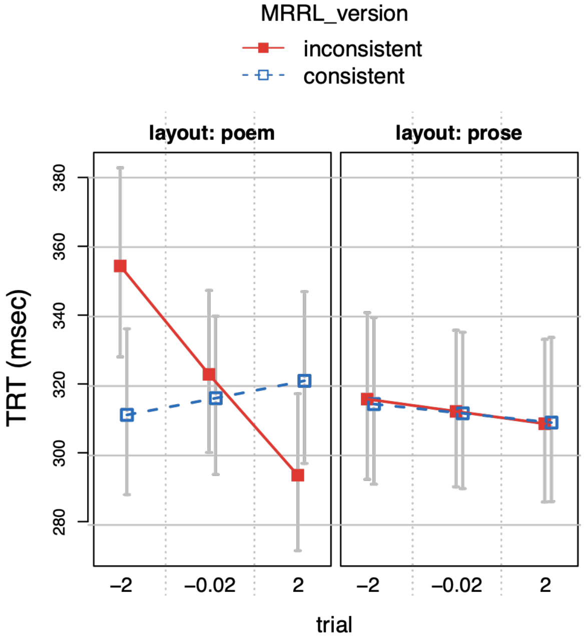

Figure 10. Total Reading Times (TRT) as a function of *layout*, *MRRL_version* and *trial* (centered). The whiskers indicate 95% confidence intervals. Raw reading times were back-transformed from the logarithmized estimates.

Verse endings (*EOV*+) showed a significant increase in all four reading time measures, indicating that verse endings were processed as *MRRL* grouping cues independent of line breaks. However, the effect was carried by the poem layout, where verse and line endings coincide (interaction *layout* by *EOV; see figure 11* for gaze durations; SFDs and RPDs show a similar pattern). This finding is partly in line with Fechino et al. (2020), who report an interaction *verse last word* by *visual presentation* for *first fixation and gaze duration*). Moreover, the presence of anomalies (*MRRL_version inconsistent*) increased *SFD*s and *TRT*s at verse endings (interaction *MRRL_version* by *EOV*).

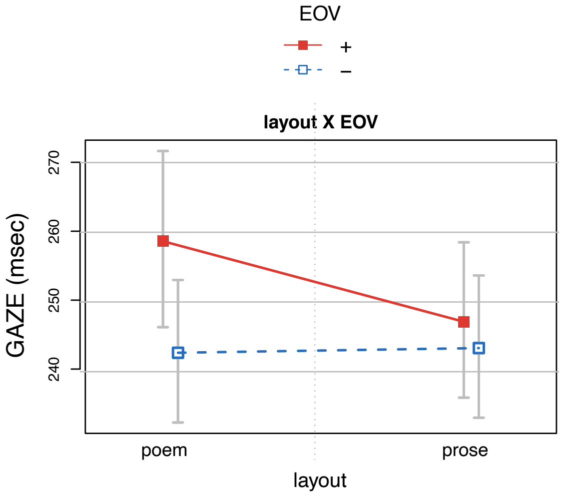

Figure 11. Gaze duration (GAZE) as a function of layout, and end-of-verse (EOV). The whiskers indicate 95% confidence intervals. Raw reading times were back-transformed from the logarithmized estimates.

Verse beginnings (*BOV*+) showed faster *SFD*s and *TRT*s, but slower *RPD*s independent of line beginnings. In *prose* layouts, *BOV*s showed increased *SFD*s, *GAZE* durations, and *TRT*s, compared to *BOV*s in *poem* layouts, where they coincide with *BOL*s. If reading times mirror pronunciation times, the effect may either indicate additional pausing for in-line *BOV*s or decreased pausing for verse beginnings at the beginning of lines. The presence of anomalies (*MRRL_version consistent*) elicited faster *TRT*s at verse beginnings. Taken together with the reversed effect at verse endings (*EOV*+), this pattern of results suggest that anomalies draw the attention to verse endings, at the expense of verse beginnings in later processing stages, as measured by *TRT*s. The *consonant vowel quotient* of the next word (*cvq.p1*) was added to establish potential parafoveal effects of pronounceability. At line endings (*EOL*+), where parafoveal preview is impossible, only GAZE durations showed significantly decreased values for higher *cvq*s on the next word (*cvq.p1*) indicating their sensitivity to potential preview effects. However, *GAZE* durations were generally smaller for higher *cvq.p1*s. *SFD*s increased with higher a *cvq.p1*, but less so in poem layouts.

Discussion

The complete model revealed that reading was slowed down in poem layout, but only in late measures (*RPD*s and *TRT*s). Increased RPDs are due to more first pass regressive saccades from the word. TRTs represent refixations to the word. Both indicate that eye progression (moving forward) decelerates, suggesting more cautious reading, and potentially maintaining a narrow eye to voice span. Furthermore, rhythmic subvocalization supposedly puts an emphasis on constantly updating local stress patterns, requiring the eyes not to jump too far ahead of the inner voice and, at the same time, force the reader into revisiting of the immediate preceding word-material. This would ultimately result in an increase of local regressive saccades, and, hence in elevated *RPD*s and *TRT*s.

*TRT*s showed interesting results in overall reading. Note that critical IAs were not included in the overall analysis (*complete model*). Reading speed increased in later trials, however mostly when anomalies were present in poem layouts. Readers were disrupted by anomalies more strongly in the poem layout at the beginning of the experiment, but got used to them in later trials, while readers basically kept the same pace throughout the entire experiment otherwise (*figure 10*).

A syllable is a single unit of speech. The complete model established the number of syllables of a word as a strong indicator of subvocalization in silent reading of poetry. Because the word length has been factored out beforehand, the residual number of syllables (*res.syllables*) represents pure pronunciation length. Hence, when word reading time increases with its number of syllables, it directly reflects the pronunciation duration of a word and is thus closely linked to subvocalization, suggesting a very narrow eye-to-(inner)-voice span.

All reading time measures were sensitive to syllables. Only for single fixation durations did word length *not* elicit a significant effect. While SFDs are usually sensitive to lexical and visual characteristics of words, these were obviously dominated by properties of pronunciation in our study. This might be due to specific task demands of our study, namely just reading MRRL-poetry without any requirement for comprehension, and seems to indicate that participants resorted to a more shallow processing mode, while they were focusing on the ‘sound of the language’ and its musical quality.

Somewhat surprisingly, the residual consonant vowel quotient *res.cvq* did not affect reading at all. The *cvq* supposedly represents the pronounceability of words, under the assumption that a higher consonant density leads to impoverished speakability. However, word-*cvq* might be just one, and possibly a minor factor of many contributing to (un-)speakability, such as slight divergencies of otherwise similar syllables in the immediate context (*Brautkleid bleibt Brautkleid*).

More importantly for this study: the *cvq* implicitly represents *syllable length*, as syllables are constructed around vowels as their nucleus, which is surrounded by a varying number of consonants. The *cvq* is thus mainly determined by the number of consonants per syllable, and therefor represents, when calculated per word, its average syllable length. Consequently, we would expect collinearities of the three variables *word_length*, *number of syllables*, and *cvq*, rendering the latter virtually redundant.

This conclusion is corroborated by our finding that in our materials, the residual number of syllables *res.syllables*, where word length is factored out, and the *cvq* were highly negatively correlated (r=-.71). Accordingly, the non-effect of *res.cvq* is no surprise. Of course, this does not mean that the *cvq* might never represent pronounceability in other materials. Whether or not it does depends on other pronunciation-related features in the particular study material captured by the *cvq*, such as sequences of consonants that are actually difficult to pronounce. In our materials however, the *cvq* did not contribute anything on top of *residual number of syllables*.

It hence remains unclear, whether the effects of the *cvq* of the successor word (*cvq.p1*) indicate preview effects of pronounceability, since we did not control for *the residual number of syllables* of the successor word. Nevertheless, the preview effect is an indicator of some lexical, and probably pronunciation related, pre-processing of the successor word. As such, it strengthens the notion of a narrow eye to (inner) voice span.

Table 6. *Complete model*. Estimates are based on logarithmized measures. Reading time measures (dependent variables) are only listed for predictors that remained in the model after stepwise elimination.

	eye.param	*Estimate*	*se*	*df*	*t*	*p*
(Intercept)	SFD	5.387	0.027	304.2	198.675	<.001
	GAZE	5.609	0.03	148.4	184.162	<.001
	RPD	6.25	0.231	22348.2	27.113	<.001
	TRT	5.778	0.037	44.2	156.429	<.001
MRRL_version inconsistent	SFD	0.006	0.003	18422.1	1.72	=.086
	GAZE	-0.004	0.003	24564.3	-1.544	
	TRT	0.004	0.005	30668.9	0.739	
layout poem	SFD	0.002	0.007	18411.5	0.322	
	GAZE	0.001	0.007	24585.4	0.173	
	RPD	0.012	0.004	24557.8	2.672	=.008
	TRT	0.017	0.006	30663.7	2.655	=.008
cvq.p1	SFD	0.008	0.003	17156.1	3.081	=.002
	GAZE	-0.008	0.004	24539.8	-2.05	=.04
res.syllables	SFD	0.027	0.007	13594.5	3.616	<.001
	GAZE	0.048	0.008	24538.8	6.233	<.001
	RPD	0.038	0.008	24550.8	4.686	<.001
	TRT	0.059	0.011	41.1	5.444	<.001
EOV +	SFD	0.022	0.004	18416.5	4.989	<.001
	GAZE	0.02	0.005	24543.9	4.081	<.001
	RPD	0.031	0.005	24553.2	6.252	<.001
	TRT	0.011	0.005	30653.5	2.048	=.041
BOV +	SFD	-0.013	0.006	18404.4	-2.039	=.041
	GAZE	-0.002	0.006	24569.2	-0.296	
	RPD	0.018	0.006	24569.5	3.078	=.002
	TRT	-0.015	0.006	30655.9	-2.628	=.009
trial	TRT	-0.012	0.003	30381.9	-4.02	<.001
word_length	GAZE	0.056	0.003	39.4	17.291	<.001
	RPD	0.065	0.003	42.2	20.539	<.001
	TRT	0.078	0.004	48.2	19.654	<.001
log.freq	SFD	-0.004	0.001	5331.1	-6.68	<.001
	GAZE	-0.009	0.001	48.9	-10.851	<.001
	RPD	-0.009	0.001	81.9	-10.165	<.001
	TRT	-0.014	0.001	30574.2	-17.575	<.001
cat C	SFD	0	0.005	11816.8	0.074	
	GAZE	0.042	0.006	24550.4	7.135	<.001
	RPD	0.053	0.006	24557.6	8.598	<.001
	TRT	0.047	0.006	30562	7.61	<.001
cat A	SFD	0.024	0.005	14829.2	4.775	<.001
	GAZE	0.005	0.005	24538.4	0.911	
	RPD	0.017	0.006	24551.5	2.999	=.003
	TRT	0.03	0.006	30390.5	5.134	<.001
cat N	SFD	-0.033	0.005	16004.5	-6.633	<.001
	GAZE	-0.04	0.005	24541.5	-7.359	<.001
	RPD	-0.056	0.006	24551.5	-9.993	<.001
	TRT	-0.069	0.006	30564.2	-11.752	<.001
EOL +	SFD	-0.075	0.005	18426.1	-13.955	<.001
	GAZE	-0.019	0.006	24548	-3.325	<.001
	RPD	-0.018	0.006	24557.7	-2.935	=.003
	TRT	-0.068	0.007	30670.7	-10.423	<.001

EOP +	SFD	0.099	0.021	18442.6	4.674	<.001
	GAZE	0.133	0.021	24554.1	6.261	<.001
	RPD	0.114	0.022	24558.4	5.184	<.001
BOL +	SFD	0.076	0.008	18352.8	9.317	<.001
	GAZE	0.084	0.008	24567.4	11.185	<.001
	RPD	0.077	0.006	24580.7	12.116	<.001
	TRT	0.027	0.007	30663.7	4.08	<.001
page	SFD	0.004	0.005	18182	0.805	
	GAZE	-0.006	0.005	24541.6	-1.079	
wpos	SFD	0.008	0.003	13499.6	3.217	=.001
	GAZE	-0.003	0.003	24570	-1.016	
	TRT	-0.02	0.003	30315.5	-6.836	<.001
gaze_pre.word	SFD	0.008	0.003	18465.1	2.601	=.009
	GAZE	-0.015	0.003	24596.1	-4.401	<.001
	RPD	-0.031	0.004	24578.5	-8.629	<.001
	TRT	-0.029	0.003	30682.1	-8.425	<.001
first_pass_regression +	SFD	-0.106	0.003	18367.8	-33.754	<.001
	GAZE	-0.169	0.004	24612.8	-47.282	<.001
	RPD	0.469	0.004	24596.2	126.433	<.001
	TRT	0.096	0.004	30688.6	24.224	<.001
MRRL_version inconsistent:layout poem	TRT	0.005	0.003	30651.7	1.587	
layout poem:cvq.p1	SFD	-0.005	0.003	18417.6	-2.035	=.042
MRRL_version inconsistent:res.syllables	SFD	0.017	0.007	18427.2	2.398	=.016
	GAZE	0.02	0.007	24570.9	2.656	=.008
MRRL_version inconsistent:EOV +	SFD	0.01	0.003	18425.6	2.883	=.004
	TRT	0.009	0.004	30662.8	2.297	=.022
layout poem:EOV +	SFD	0.022	0.004	18431.9	4.972	<.001
	GAZE	0.012	0.005	24571	2.554	=.011
	RPD	0.013	0.005	24558.8	2.595	=.009
	TRT	0.027	0.005	29965.4	5.151	<.001
MRRL_version inconsistent:BOV +	TRT	-0.01	0.004	30665	-2.411	=.016
layout poem:BOV +	SFD	-0.023	0.006	18446.7	-3.752	<.001
	GAZE	-0.012	0.006	24570	-1.979	=.048
	TRT	-0.015	0.005	30647.2	-2.899	=.004
MRRL_version inconsistent:trial	TRT	-0.014	0.003	30577.5	-4.303	<.001
layout poem:trial	TRT	-0.007	0.003	30205.9	-2.242	=.025
cvq.p1:EOL +	GAZE	-0.008	0.004	24546.5	-2.065	=.039
page:wpos	SFD	-0.009	0.004	18112.7	-2.244	=.025
	GAZE	-0.015	0.005	24559.3	-3.291	=.001
MRRL_version inconsistent:layout poem:trial	TRT	-0.013	0.003	29890.3	-4.053	<.001

Note. For SFD, the number of observation was 18505, the conditional R^2 was 0.159 and the marginal R^2 was 0.092. For GAZE, the number of observation was 24659, the conditional R^2 was 0.242 and the marginal R^2 was 0.172. For RPD, the number of observation was 24638, the conditional R^2 was 0.51 and the marginal R^2 was 0.448. For TRT, the number of observation was 30765, the conditional R^2 was 0.279 and the marginal R^2 was 0.149.

EOV + stands for *end-of-verse* = true, BOV stands for *beginning of verse*, EOL stands for *end of line*, BOL stands for *beginning of line, wpos* represents the *serial order position of a word on a page.* For further information, see section *predictors* in *data analysis. Model specification: log.<duration> ~ MRRL_version * layout * res.cvq + MRRL_version * layout * cvq.p1 + MRRL_version * layout * res.syllables + MRRL_version * layout * EOV + MRRL_version * layout * BOV + MRRL_version * layout * trial + word_length + log.freq + res.syllables + res.cvq + cvq.p1 + cat + EOL + cvq.p1:EOL + EOP + BOL + page * wpos + gaze_pre.word + trial + first_pass_regression + (1 | vp) + (0 + res.syllables + word_length + log.freq | vp) + (1 | item)*

Skipping probability. Across all conditions, words were skipped with an average rate of .24 (*sd* = 0.16). The most frequently skipped words were short function words and pronouns at the beginning of a new line (e.g., *Er*, *Am*, *So*, *Da*), with a skipping probability up to .88. Twenty words were never skipped at all (e.g., *aufgezehrt*, *Chorusgleis*, *denkt*, *Feuertopf*).

For skipping probability, we did not fit the main model for the critical IAs, because almost no skipping was found here. However, we fitted the complete model. Due to the binary nature of the response variable, we fitted a logistic regression using a similar predictor structure (excluding *cvq.p1* and including the interactions *word.length:BOL*, *log.freq:BOL*, *cat:BOL*, *layout:BOL, page:layout*, *wpos:layout*) of the complete model in earlier fits.

Table 7. Skipping probability.

	Estimate	*se*	*z*	*p*
(Intercept)	-0.599	0.198	-3.025	=.002
MRRL_version inconsistent	-0.004	0.023	-0.162	
layout poem	-0.066	0.030	-2.173	=.03
res.cvq	-0.075	0.018	-4.260	<.001
res.syllables	0.079	0.040	1.971	=.049
EOV +	0.091	0.026	3.542	<.001
BOV +	0.044	0.162	0.271	
trial	0.059	0.013	4.560	<.001
word_length	-0.386	0.016	-24.477	<.001
log.freq	0.005	0.006	0.870	
cat C	0.243	0.041	5.939	<.001
cat A	-0.041	0.038	-1.068	
cat N	-0.135	0.051	-2.658	=.008
EOL +	-0.350	0.031	-11.297	<.001
EOP +	-0.043	0.156	-0.273	
BOL +	0.809	0.163	4.968	<.001
page	0.144	0.024	6.117	<.001
wpos	-0.020	0.013	-1.500	
gaze_pre.word	-0.092	0.014	-6.446	<.001
MRRL_version inconsistent:layout poem	-0.027	0.023	-1.200	
MRRL_version inconsistent:res.cvq	-0.011	0.017	-0.634	
layout poem:res.cvq	-0.015	0.017	-0.867	
MRRL_version inconsistent:res.syllables	0.006	0.038	0.160	
layout poem:res.syllables	0.068	0.038	1.775	=.076
MRRL_version inconsistent:EOV +	-0.017	0.019	-0.868	
layout poem:EOV +	0.131	0.025	5.240	<.001
MRRL_version inconsistent:BOV +	-0.015	0.016	-0.933	
layout poem:BOV +	-0.221	0.162	-1.366	
MRRL_version inconsistent:trial	0.005	0.014	0.385	
layout poem:trial	0.022	0.014	1.565	
page:wpos	-0.013	0.023	-0.590	
word_length:BOL +	-0.095	0.016	-6.084	<.001
log.freq:BOL +	-0.015	0.006	-2.559	=.01
cat C:BOL +	0.120	0.040	2.976	=.003
cat A:BOL +	0.099	0.038	2.614	=.009

cat N:BOL +	-0.221	0.050	-4.406	<.001
layout poem:BOL +	0.086	0.162	0.531	
layout poem:page	0.060	0.024	2.529	=.011
layout poem:wpos	0.049	0.013	3.661	<.001
MRRL_version inconsistent:layout poem:res.cvq	0.002	0.017	0.096	
MRRL_version inconsistent:layout poem:res.syllables	-0.077	0.038	-2.027	=.043
MRRL_version inconsistent:layout poem:EOV +	-0.036	0.019	-1.892	=.058
MRRL_version inconsistent:layout poem:BOV +	0.018	0.016	1.126	
MRRL_version inconsistent:layout poem:trial	0.041	0.014	2.937	=.003

Note. For skipping probability, the number of observation was 37221, the conditional R^2 was 0.283 and the marginal R^2 was 0.0885. EOV + stands for *end-of-verse* = true, BOV stands for *beginning of verse*, EOL stands for *end of line*, BOL stands for *beginning of line, wpos* represents the *serial order position of a word on a page.* For further information, see section *predictors* in *data analysis. Model specification: skip ~ MRRL_version * layout * res.cvq + MRRL_version * layout * res.syllables + MRRL_version * layout * EOV + MRRL_version * layout * BOV + MRRL_version * layout * trial + word_length + log.freq + res.syllables + res.cvq + cat + EOL + EOP + BOL + page * wpos + gaze_pre.word + trial + word_length:BOL + log.freq:BOL + cat:BOL + layout:BOL + page:layout + wpos:layout + (1 | vp) + (1 | item)*

Generic variables. The logistic mixed effects regression fit (*table 7*) shows that many variables representing lexical or structural features of words behaved as expected. Skipping probability drastically decreased with word length (*word_length*), while word frequency only affected skipping after the line initial word (*log.freq* by BOL interaction). Skipping was thus generally strongly influenced by shape related visual features, whereas word recognition and lexical access played a role only in parafoveal preview. The reliable effect for *page* shows that skipping increased on later pages of a trial. Skipping was also increased in *poem* layouts (main effect *layout*), particularly towards the end of single pages as well as on later pages, towards the end of poems (*layout poem:wpos*, *layout poem:page*). It also increased with later trials (*trial*) and even stronger when anomalies were present in layout poem (*layout poem:MRRL_version inconsistent:trial*).

Words were skipped significantly more often at beginnings of lines (main effect *BOL+*). The amount of skipping of line initial words was modulated by a variety of lexical variables though: first and foremost, line initial words were skipped much more often when they were very short (*figure 12*).

Word category affected skipping as a main effect, but also had a strong impact on skipping the first word of a line (*cat:BOL*). As illustrated in figure 13, nouns were the least likely category to be skipped here, followed by verbs and adjectives. Closed-class words (*cat C*) were skipped most often. Note that this category effect cannot be attributed to the fact that closed class words are short and open class words (*N*, *V*, and adjectives) are longer on average, because *word_length* was taken into account independently, as were other lexical variables.

Interestingly, layout poem did not influence skipping at line beginnings (*layout:BOL*) in addition to the aforementioned variables.

Among the variables related to pronunciation, a higher *consonant vowel quotient* (*res.cvq*) significantly diminished the skipping probability for a word. In layout prose, the residual number of syllables significantly diminished skipping for consistent versions (*MRRL_version:layout:res.syllables*).

While skipping was less likely at the end of lines (*EOL+*), it increased significantly at the end of verses (*EOV*), even more so in the poem layout (*layout poem:EOV*), and marginally so when anomalies were included (*MRRL_version inconsistent:layout poem:EOV*). Verse endings notably coincide with line endings in poem layout, but – with few exceptions – did not so in prose layout. As verse endings are the place where anomalies were realized, it stands to reason that *EOV*s attract more attention in the *inconsistent MRRL_version*.

Skipping probability decreased with larger gaze durations on the previous word (*gaze_pre.word*) suggesting that if the previous word was hard to process, parafoveal preview of the following word might have been limited so that skipping became less likely.

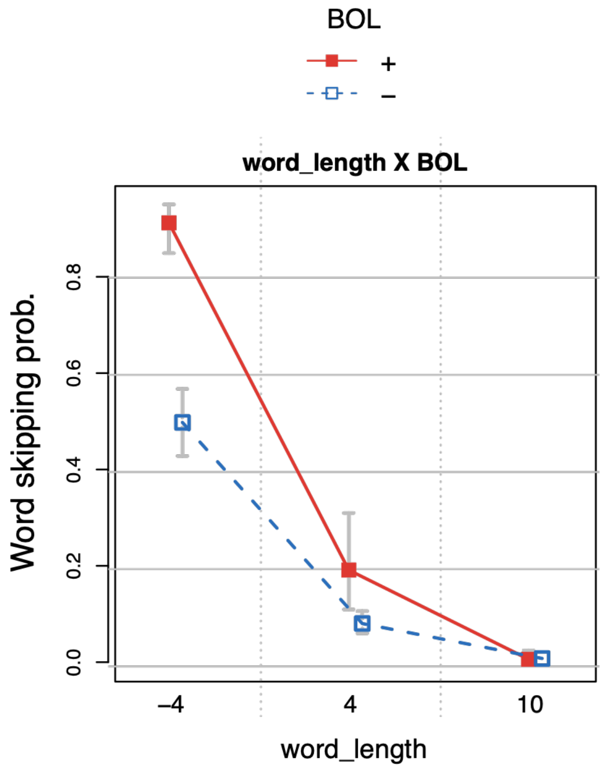

Figure 12. Word skipping probability as a function of word length (*word_length*) and beginning of line (BOL). The whiskers indicate 95% confidence intervals. Word length was centered hence the values range from -4 to 10 with a mean of 4.

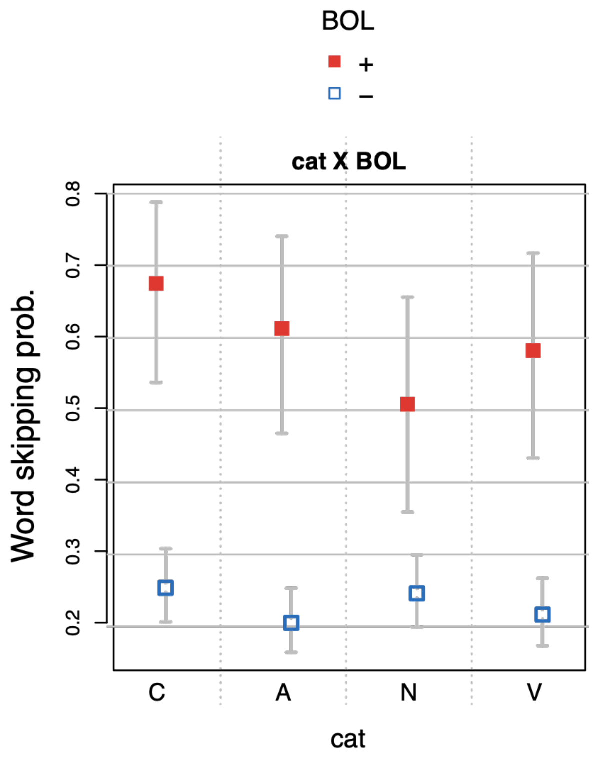

Figure 13. Word skipping probability as a function of word category (*cat*) and beginning of line (*BOL*). The whiskers indicate 95% confidence intervals.

Discussion

Word-skipping mainly showed expected results with respect to lexical and formal features indicating that skipping in MRRL often mirrors the results of reading times, but also diverges in some respects.

Words at verse ends were skipped more often, and more so in poem layouts.

The presence of anomalies (*MRRL_version inconsistent*) in *poem layout* items, however, sligthly reduced (p=.058) skipping at the end of verses (*EOV+*). Since anomalies always occurred at the end of both verses and lines in *poem layout*, their presence had apparently induced a more cautious reading style at this particular position.

Earlier, we argued that in our study, the *cvq* is highly confounded with average syllable length, which explains the lack of a *cvq* effect on top of a highly reliable effect of residual syllable number in reading times. Nevertheless, we included *res.cvq* in the skipping model because *skipping* might still be more sensitive to pronounceability. In fact, the *res.cvq* significantly decreased skipping, while the residual number of syllables was inconclusive. The differential effects of *res.syllables* and *rec.cvq* make sense though. As syllables are the units of speech, their number should directly affect pronunciation *time*. The probability of a skipping event, however, strongly depends on the word's length and frequency, i.e., measures of its recognizability and lexical accessibility. Syllable-based pronunciation *duration* might thus add no valuable skipping criterion, whereas pronounceability might very well do. The data support this assumption: skipping was significantly reduced with elevated consonant densities, i.e., higher *res.cvq*s, whereas the number of syllables had virtually no effect on over-all skipping.

The three-way interaction of layout, MRRL-version and syllables, however, mirrors the fact that skipping probability shrinks with the number of syllables only in prose layout when no anomalies were present. This finding suggests that reading in prose layout might have been more cautious and more closely aligned with the inner voice, as long a reading was not disrupted by anomalies.

At line beginnings, *word-length* and word *category* strongly affected skipping, whereas *layout* did not. This finding does not appear to support (Blohm, 2020) finding that poem layouts induce more cautious reading. However, he reports the strongest effect for the first function word in text medial position in poem vs. prose layouts. Our BOL effect is mainly driven by word-length and category, with nouns being the least likely words to be skipped. This might appear surprising at first glance, as the category of the line initial word cannot be processed in parafoveal preview from the final words of the previous line. However, word category can often be predicted from the preceding syntactic context. Note that syntactic boundaries are also syntactic prediction cues. Sentences typically start with a determiner or (short) function words such as *and, it,* etc*.,* in particular in our stimuli. Additionally, predictable beginnings in poems are often used as a rhetoric figure of *repetitio*. Hence, readers may be able to statistically predict in a poem, which word may start the beginning of a next line. Line initial word skipping thus appears to depend upon both, bottom-up perceptual features as captured by word length and top-down prediction-based information, such as word category.

Load contributions of pre-rhymes

The LC measure calculates selective re-reading and allows an investigation of the time spent re-reading (sum of all fixation durations on) a previous region in the regression path of a later region. It helps to indicate whether the eyes re-fixate the pre-rhymes when a regressive saccade is triggered by a rhyme anomaly. We analyzed an *IA subset* containing all rhyme words, i.e., the last word of the third (A) and fourth verse (B) in each stanza, and computed the time spent on the corresponding pre-rhyme word, i.e., the last word of the first or second verse respectively, when a regressive saccade was launched from the rhyme word after first pass reading.

The variable *anomaly_type* has been coded only for critical interest areas, which introduced anomalies in the inconsistent MRRL_version. All other rhyme words at non-critical positions were coded *zero*, indicating that the wording was identical for consistent and inconsistent MRRL-versions (*see figure 14*). This allowed us to analyze the effect of rhyme and metric anomalies on the load contributions of pre-rhymes in all stanzas. This *zero* condition, which did not contain any anomalies, served as a baseline condition.

The model fit (table 8) shows that the time spent on a pre-rhyme within a regression path of the corresponding target IA increased significantly when both *metric* and *rhyme* were manipulated in the poem layout (*see figure 14*). The effect amounts to a robust significant three-way interaction of factors *layout*, *MRRL_version* and *anomaly_type r&m* (*anomaly_type r&m:layout poem:MRRL_version inconsistent*), on top of the two-way interactions of *layout poem* and *anomaly_type rhyme and meter* (*anomaly_type r&m* x *layout),* and *layout poem* and *anomaly_type rhyme* (*anomaly_type rhyme:layout poem*), as well as on top of the two-way interaction of *anomaly_type rhyme:MRRL_version inconsistent*, and two significant main effects, one for *r&m* anomalies (*anomaly_type r&m*), and one for *layout poem* (*layout poem*).

For anomaly type *r&m* in *poem* layouts, we found a significant post-hoc contrast between the inconsistent and consistent MRRL_version (*est* = *0.6067, se = 0.220, df = 463, t-ratio = 2.757, p=.006*). For anomaly type rhyme, post-hoc contrast was significant between the inconsistent and consistent MRRL_version in *poem layout* (*est = 0.5771, se = 0.315, df = 632, t-ratio = 1.832, p=.067*), as well as marginally so in *prose layout* (*est = 0.5937, se = 0.311, df = 644, t-ratio = 1.911, p=.057*).

Table 8. Re-reading time on pre-rhyme within regression path of critical IA.

	Estimate	*se*	*df*	*t*	*p*
(Intercept)	0.712	0.189	505.8	3.775	<.001
anomaly_type metric	-0.086	0.124	62.2	-0.696	
anomaly_type r&m	0.429	0.117	105.5	3.651	<.001
anomaly_type rhyme	-0.109	0.114	52.3	-0.964	
layout poem	0.264	0.087	58.8	3.046	=.003
MRRL_version inconsistent	0.064	0.057	59.8	1.126	
wpos	0.264	0.142	1149.8	1.859	=.063
page	-0.204	0.105	1158.6	-1.938	=.053
trial	0.001	0.042	910.2	0.015	
anomaly_type metric:layout poem	0.010	0.115	1106.2	0.083	
anomaly_type r&m:layout poem	0.301	0.075	1171.4	4.031	<.001
anomaly_type rhyme:layout poem	-0.235	0.090	1074.8	-2.615	=.009
anomaly_type metric:MRRL_version inconsistent	-0.177	0.096	1032.8	-1.840	=.066
anomaly_type r&m:MRRL_version inconsistent	0.026	0.069	1136.4	0.382	
anomaly_type rhyme:MRRL_version inconsistent	0.229	0.088	1065.5	2.610	=.009
layout poem:MRRL_version inconsistent	0.029	0.046	1160.8	0.624	
wpos:page	-0.181	0.087	1156.8	-2.075	=.038
anomaly_type metric:layout poem:MRRL_version inconsistent	-0.071	0.097	1007.1	-0.737	
anomaly_type r&m:layout poem:MRRL_version inconsistent	0.184	0.069	1156.9	2.658	=.008
anomaly_type rhyme:layout poem:MRRL_version inconsistent	-0.033	0.087	1137.5	-0.380	

Note. The number of observation was 1243, the conditional R^2 was 0.288 and marginal R^2 was 0.076.
Model specification: lmer Table: log(lc_pre + 1) ~ anomaly_type * layout * MRRL_version + wpos * page + trial + (anomaly_type + layout + MRRL_version | vp) + (1 | item)

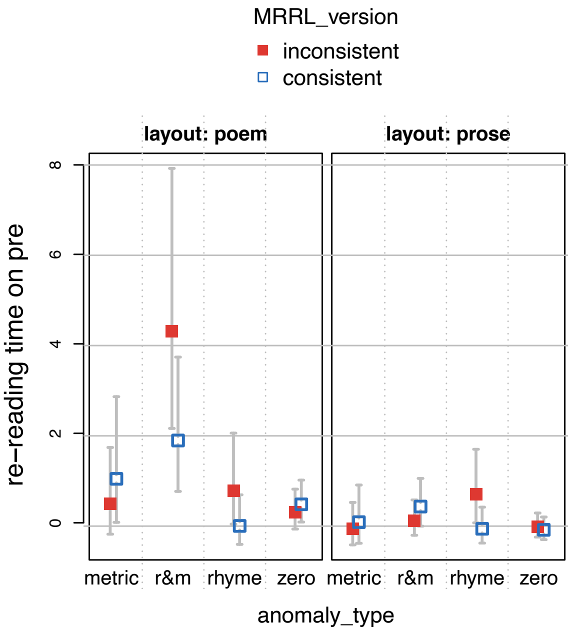

Figure 14. Re-reading time (msec) on prime as a function of *layout*, *version* and *anomaly type*.

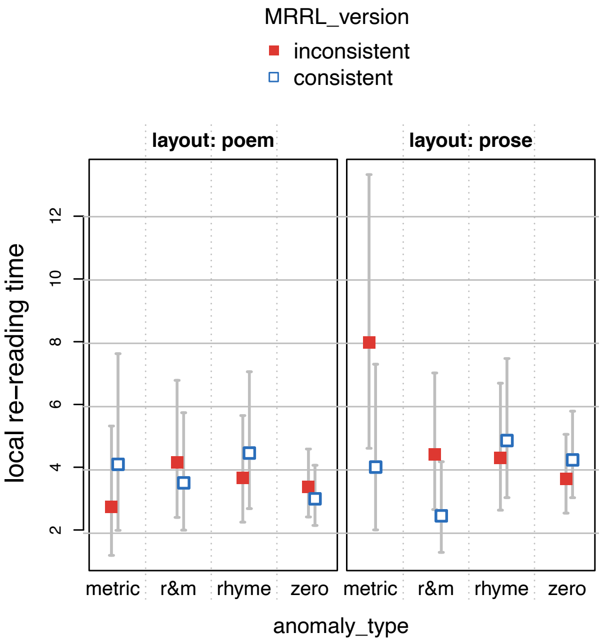

Figure 15. Re-reading time (msec) on *local context* (one to six words before critical IA) as a function of *layout*, *version* and *anomaly type*

We also calculated *load contribution* for the material directly preceding the anomaly. We had predicted that metric anomalies elicit more re-reading the verse that it occurs in, as this is where its rhythmic gestalt is established. This hypothesis seems to be partly supported. The model fit (see table 9, *figure 15*) shows that the time spent on the *local context* (one to six words before critical IA), after a first pass regression was launched from the critical IA, yielded two-way interactions *anomaly_type r&m* x *layout*, and *anomaly_type r&m x MRRL_version,* as well as a three-way interaction *anomaly_type metric* x *layout* x *MRRL_version*.

For anomaly type *metric* in *prose* layouts, we found a significant post-hoc contrast between the inconsistent and consistent MRRL_version (*est = 0.573, se = 0.241, df = 973, t-ratio = 2.375, p=.0178*). For anomaly type *r&m*, post-hoc contrasts were significant between the inconsistent and consistent MRRL_version, too (*est = 0.437, se = 0.177, df = 458, t-ratio = 2.472, p=.014*). However, metric anomalies did not induce local re-reading in *poem* layouts as hypothesized.

Table 9. Re-reading time (msec) on *local context* (one to six words before critical IA)

	Estimate	*se*	*df*	*t*	*p*
(Intercept)	1.352	0.168	178.9	8.049	<.001
anomaly_type metric	0.090	0.106	39.6	0.850	
anomaly_type r&m	-0.076	0.103	80.2	-0.739	
anomaly_type rhyme	0.068	0.085	47.8	0.803	
layout poem	-0.071	0.042	472.1	-1.700	=.09
MRRL_version inconsistent	0.035	0.048	43.6	0.718	
wpos	-0.448	0.107	5562.4	-4.183	<.001
page	0.169	0.079	5301.5	2.129	=.033
trial	0.010	0.032	2190.4	0.327	
anomaly_type metric:layout poem	-0.138	0.088	3578.6	-1.576	
anomaly_type r&m:layout poem	0.124	0.056	6815.5	2.197	=.028
anomaly_type rhyme:layout poem	0.023	0.067	3918.2	0.343	
anomaly_type metric:MRRL_version inconsistent	0.033	0.073	2876.4	0.456	
anomaly_type r&m:MRRL_version inconsistent	0.107	0.052	5859.9	2.049	=.041
anomaly_type rhyme:MRRL_version inconsistent	-0.098	0.065	3429.6	-1.496	
layout poem:MRRL_version inconsistent	-0.064	0.035	5320.3	-1.828	=.068
wpos:page	0.185	0.066	7200.4	2.828	=.005
anomaly_type metric:layout poem:MRRL_version inconsistent	-0.154	0.073	2583.3	-2.100	=.036
anomaly_type r&m:layout poem:MRRL_version inconsistent	-0.012	0.052	6419.1	-0.235	
anomaly_type rhyme:layout poem:MRRL_version inconsistent	0.050	0.065	5027.6	0.776	

Note. The number of observation was 7458, the conditional R^2 was 0.088 and marginal R^2 was 0.008.
*Model Specification: log(lc_local + 1) ~ anomaly_type * layout * MRRL_version + wpos * page + trial + (anomaly_type + layout + MRRL_version | vp) + (1 | item)*

Discussion

The raw *load contribution* measure was introduced to account for selective re-reading.

Re-reading of pre-rhyme. We found that *rhyme* anomalies in both the *poem* and the *prose layout* induced re-reading of the pre-rhyme. Readers appear to utilize pre-rhymes for resolving the rhyme anomaly across lines.

For metrical anomalies it seems plausible that refixation of the pre-rhyme is not necessary for resolving metrical anomalies, when the expected rhyme scheme is met.

Strikingly, combined meter and rhyme anomalies (*r&m*) in the poem layout elicited the highest amount of re-reading of the pre-rhyme. The fact that rhyme-meter anomalies (*r&m*) only elicited re-reading in the *poem layout* suggests that, in prose layout, the combined absence of metrical, rhyme and visual layout cues appeared to leave readers disoriented, whereas readers could at least use the visual cues of strict poem layouts to check for the pre-rhyme in case of a severe anomaly. This, of course, requires that the anomalies had been noticed beforehand.

In general, the results indicate that readers were able to process the rhyme scheme, which they only could have accomplished by processing and representing phonetic and phonological information.

Re-reading of local context. For *metric* (and *r&m*) anomalies we expected more re-reading of the material preceding the anomaly in the current verse because it establishes the metrical pattern leading to the last word where the metric is broken. This hypothesis was supported for prose but not for poem layouts. This somewhat surprising interaction with layout may be due to the fact that re-reading the current verse in prose layout required jumping back to the previous line in some stimuli, whereas they could stay in the current line in poem layout. In our analysis, we operationalized local context as the six words preceding an anomaly. This window may have exceeded the number of words in the line though. Importantly, we found only rare re-visits of words farther than three words away. This suggests that readers remained within the current verse, and that words before it had no big impact on our results.

In summary, our findings on re-reading support the assumption that readers are sensitive to different types of rhythmic expectation violations. They react differently to metric than to rhyme anomalies, with vastly different re-fixation patterns.

General Discussion

Our goal was to investigate silent MRRL reading. We hypothesized that readers would pick up the rhythmic patterns induced by the sound of words and their phonological stress pattern in metrically regular and rhymed language. We found a multitude of indicators for rhythmic subvocalization. Readers responded to metrical anomalies with longer reading times both on the anomalies themselves and by re-reading the preceding material. Similarly, rhyme, as well as *rhyme&meter* anomalies, were read longer and triggered systematic re-reading of pre-rhymes, particularly in the poem layout. This finding suggests that the metric and rhyme structure of the stanzas had been picked up very well. The differential effects for the anomaly types in different layouts, however, deserves some closer examination.

Increased fixation and reading times for *metric* anomalies were found for all four reading time measures, albeit only in the poem layout. Here, readers clearly seemed to be disrupted when the metrical grid was violated. The fact that readers appeared to stumble less over metrical anomalies in prose layouts is somewhat puzzling. Apparently, the overlay of line breaks and verse endings is crucial for establishing a verse's metrical grid. This may have two possible reasons: Firstly, when lines are composed of two distinct verse segments, which in our stimuli was often the case in prose layout, they may have been processed as enjambments, as investigated by Koops van't Jagt et al. (2014), or simply as metrically disconnected. In the latter case, metrical processing may be impeded. This assumption is supported by the lack of metrical anomaly effects in our reading time data. In the former case, when the verse structure is experienced as intact even though it spans over two lines, the metrical structure should have been experienced as well. This assumption is corroborated by the finding of increased selective re-reading of the preceding words in prose layouts when regressive saccades were prompted by metric anomalies after first pass reading. These cases, albeit rare, suggest that metric anomalies in prose layouts do bear the potential to elicit specific eye movement responses.

Secondly, metrical anomalies may be more tolerable in prose layouts as long as the rhyme is intact. In the absence of visual poetic cues, the poetic form may just have been experienced as less strict.

Rhyme anomalies, on the other hand, caused stronger reading time effects in the *prose* layout in gaze durations and regression path durations. In the *poem* layout, *rhyme* anomalies show no such effect and appear to merely elicit an adoption of a different rhyme scheme, such as ABAC.

One explanation for the interaction with layout could be that, in the absence of a clear poetic visual form in *prose* layout, rhymes may serve as "poetic anchors" to support the structural expectations induced by the MRRL environment. In *poem* layout, visual cues and other stylistic devices may serve the same purpose, so that no such anchor is necessary. Therefore, rhyme violations become more tolerable, and readers easily accommodate rhyme deviations into a slightly diverging rhyme scheme.

This is in line with the notion that rhyme has the potential to end stanzas and poems as a rhythmic unit by eliciting a "sense of closure" (Smith, 1968), as a phenomenological experience. In the *poem* layout, a sense of closure can be induced by the visual 'gestalt' of a stanza/poem. The *function of closure* account is supported by Fechino et al. (2020), who report overall longer first fixation durations, gaze durations and total reading times for rhyme words when presented in verse layout. They also report increased rereading probabilities for rhyme words in the prose layout, indicating that readers have processed the rhyme. In our consistent versions, however, we did not find this contrast.

Menninghaus and Wallot (2021) reported longer total word reading times (total gaze durations) for poem with higher appreciation scores. At the same time, they found that rhyme anomalies had the strongest negative effect on appreciation, whereas metric anomalies reduced appreciation about half as much (but see Skov and Nadal, 2020, for a critique of the construct of "aesthetic emotions"). Could the reading time effects of anomalies and layout in our study hence be caused by differential effects of appreciation? The pattern of our results does not suggest so. Firstly, rhyme anomalies did not elicit stronger reading time increases than metric violations, it is rather the other way round. Secondly, both effects interacted with layout in different ways. Although we do not have data on how layout affects appreciation, we would suspect that the poem layout would receive higher appreciation scores than prose. If appreciation was the mediating variable, we would expect the shortest total word reading times for *rhyme* – and *r&m* – anomalies in *prose* layout, and the longest for *consistent* versions in *poem* layout, with *rhyme* anomalies in *poem* layouts, *metric* anomalies, and *consistent* version in *prose* layout somewhere in between. This is not what we found. Instead, our pattern of results suggest that readers' reactions were due to specific processes of accommodation and repair caused by the very nature of the anomalies and the layout.

Scheepers et al. (2013) found strong effects of rhyme anomalies in listeners' pupillary responses, but the effects of metrical and other anomalies were much smaller. In our study we found the strongest effect on reading times for metrical anomalies. The two studies used very different dependent measures, as well as different stimuli types, and, most importantly, different presentation modalities. Spoken stimuli, as in Scheepers et al.'s study, provide all phonological cues necessary to directly perceive a rhythmic gestalt, whereas readers are forced to reconstruct the auditive gestalt from subvocalization of visual stimuli.

We assume that subvocalization is cognitively more demanding than listening to external speech, as subvocalization entails language production on top of visual language processing. If this is the case, the processing of rhymes spanning two lines may be particularly more demanding in silent reading than in listening. Metric processing however can be achieved via subvocalization of only a few local words and may therefore be less demanding, hence the different results. Moreover, pupillary responses in Scheepers et al.´s study probably reflect processes on an affective or aesthetic level, where rhymes or rhyme violations may trigger stronger responses.

Previous research states that readers generally adjust their reading style and pace to the text genre and that the poem layout is a relevant cue for such an adjustments (Blohm et al., 2017; Hanauer, 1998b, 2001; Menninghaus & Wallot, 2021; Peskin, 2007; Xue et al., 2020). In our study, the rhyme effects in gaze durations and RPDs suggest that MRRL is detectable without the layout cue. Hence, layout is neither necessary nor sufficient for identifying a MRRL-text as poetry. This is in line with conclusions drawn in Fechino et al' (2020, p. 13).

We also analyzed how different anomaly types triggered readers to re-read certain portions of text systematically. Re-reading time (load contribution) of the pre-rhyme revealed that readers captured the rhyme scheme in both layouts and used it for cross-line re-orientation towards the pre-rhyme when deviations from the overall dominant rhyme pattern in MRRL occurred. Interestingly, re-reading of pre-rhymes was activated strongest for rhyme-meter anomalies in layout poem, whereas no such effect was found for the prose layout.

Load contribution appear to be a sensitive measure in this specific case, since the results clearly suggest that readers did capture a combined rhyme and meter anomaly in poem layouts as an anomaly, thus experiencing a sense of violation of the overall metrical grid. However, the visual-spatial cues of a strict poem layout seem to somehow facilitate orientation towards the pre-rhyme. Taken together, the effects of RPD and LC suggest that the rhyme scheme had been picked up and processed by readers in both the poem and the prose layout. However, the visual cues of a poem layout made it easier to resolve severe rhyme and meter anomalies.

We also included two variables closely linked to pronunciation as indicators of subvocalization: *number of syllables* and the *consonant vowel quotient, cvq*, both residualized for word length (the *cvq* was also residualized for residual number of syllables, to keep both variables independent).

Previously, the *number of* s*yllables* in a word has been reported to have no additional effect on reading times beyond *word length* in normal reading (Fitzsimmons & Drieghe, 2011). However, in our experiment on MRRL-reading, we found that all reading times were highly sensitive to syllables, on top of effects of word length and frequency, indicating that silent reading of MRRL is closely tied to pronunciation. In fact, SFDs only showed a clear syllable effect but no effect of word length, and thus appears to be linked more closely to pronunciation in our study. Crucially, the syllable effect was increased in text versions with anomalies in early processing measures (SFD and GAZE). This result in particular indicates that readers resorted to even more intense subvocalization when reading gets disturbed by MRRL anomalies, presumably narrowing the eye-(inner)-voice span even further.

Contrary to the assumption that a higher consonant density should reduce speakability and thus slow down subvocalization, the residual consonant vowel quotient did not affect reading times. We found that in our material, *cvq*s were highly correlated with residual syllable length (where word length had been cancelled out). Nevertheless, *cvq*s turned out to be a much better predictor for word skipping, where the number of syllables played virtually no role. Taken together, the findings suggest that fixation duration measures were strongly influenced by pronunciation *duration,* as approximated by syllable number, whereas skipping is more affected by *pronounceability* as approximated by the residual *cvq*.

At line beginnings, both *word-length* and word *category* strongly affected skipping. Line initial word skipping thus appears to depend upon both bottom-up perceptual features, such as shape, and top-down prediction-based information, such as word category. Presentation *layout,* however, did not have an additional effect. This result stands in contrast to the assumption that the poem layout itself induces more cautious reading at line beginnings (Blohm, 2020). There was a notable difference between the materials of the two studies though: in the Blohm study, all lines in the poetry layout started with a capital letter (ibid. 72), whereas ours did not. Hence, capitalization might have attracted attention to line initial words, rather than a more cautious reading style.

We conclude that our finding indicates a high grade of synchronization of the eyes with inner speech (see Silva, Reis, Casaca, Petersson, & Faísca, 2016 for a discussion on the topic of voice-eye-lead), induced by the rhythmic structure of MRRL-language.

We would like to add some general remarks on subvocalization. There might be several levels of representing an “inner -rhythmic - voice” while reading MRRL silently: Ranging from an abstract representation of implicit prosody (compare Breen, 2014) to the specific, conscious use of an inner voice (Alderson-Day et al., 2017). Also, automatized – yet vulnerable – vs. controlled rhythmic processing may play a role. However, for all ‘levels’, we presume that for rhythmic subvocalization of MRRL, phonological awareness (Cason & Schön, 2012; Melby-Lervåg et al., 2012) is crucial for the detection of ‘the beats between and on the sounds’ (Langus et al., 2017; Tierney et al., 2017). Such a process, to us, may be closely related to the ability to induce a beat as well as to the potential to get entrained.

However, this first investigation of subvocalization in MRRL using eye-tracking measures can only shed initial light on the topic. Eye-movements may not bear a direct index of entrainment processes - but supposedly an additional one (compare Lange et al., 2018; Nozaradan, 2014). That said, we would like to point out that other evidence for rhythm representation and processing comes from studies on musical notation reading, indicating a temporal, melodic and pause-bound representation of rhythm as reflected in eye movements (Silva & Castro, 2019).

Also, unlike other experiments which focused particularly on the interaction of metric anomalies and sentence processing (Breen & Clifton, 2011, 2013), our study had a somewhat broader starting point. The perception and processing of anomalies is just one factor among a variety of variables we looked at, among e.g., the layout variation and others. Most importantly, we were interested in MRRL induced reading styles *at other positions than the critical interest areas* to better understand how MRRL elicits rhythmic subvocalization and induces entrainment. Consequently, we had to use longer stimuli to be able to compare possible interactions. We also deliberately chose to not include comprehension or memory tasks (Tillmann & Jay Dowling, 2007; Xue et al., 2020), because both tasks might have changed the reading style with respect to rhythmic subvocalization. We do presume that the processing of a poem's rhythm can both guide and impede comprehension, however, this question was not within the scope of our study.

Another important aspect to discuss is the nature of the stimulus material itself. The poems were constructed in a way such that virtually no additional pausing, lengthening or shortening of syllables was required to realize a relatively uniform rhythm. The syllable structure thus aligned easily with the abstract metrical grid. We therefore think that a (linguistic) conceptual difference between meter and rhythm can be neglected in our study.

It also remains an open question, whether a reader's subvocalization realizes the full rhythmic scope of MRRL or merely so in degrees. For example, if one is not experienced in expressive reciting, the prosody of 'daily speech' may shape subvocalization and comprise it in silent reading MRRL to a more binary stance, i.e., strong-weak stressing. Then, silent reading may be aligned solely with a 'tactus', i.e., the induced beat. Hence, subvocalization would not necessarily imply strong phrasal, intonational or modulating aspects.

However, given the assumption that this type of MRRL can produce a reasonably stable periodicity, i.e., 'quasi-isochrony' (Patel, 2003, 2010), the following should hold: „all isochronous sequences are rhythmic, but not vice versa" (Ravignani & Madison, 2017). That said, we propose that MRRL eliciting rhythmic subvocalization - which we found strong indicators for – may linger between idealized vs. empirical isochrony. In *silent* reading MRRL, related cognitive processes may not necessarily be provoked by the physical signal itself, but they are most certainly based on the „psychological tendency to superimpose an isochronous grid to a rhythmic sequence" (Ravignani & Madison, 2017).

Limitations

First, reading experiments using natural and, in particular, long texts are not only rare but also prone to error. Half of our stimuli, which were written specifically for the purpose of the experiment, included "anomalies" as described in the Design and Materials section. A rhyme-scheme error was spotted within the first stanza of "9 Leben". Here, the ABAB-scheme was not met ("Morgen" in line 3, instead of "Lande"). However, we do believe that this did not carry weight for beat extraction and expectation of a rhythmic figure throughout the rest of the poem. Two minor corrections had to be made for two words in the stimuli during data collection. There was a mistake at the last word of stanza 5 (not a critical IA) in the modified version of item "Flüstern", where the word „wertig" had to be replaced with "fein", as in the original version. This mistake was corrected from participant nr. 23 on. Another mistake was a small number of wrong apostrophe symbols and one orthographic mistake. This was corrected from participant nr. 29 on. For data analysis, the affected data points on the two IAs were coded accordingly and did not affect the overall results.

Secondly, one can criticize that the anomalies chosen have not been consistent with respect to the *type* of rhythmic deviation, e.g., adding one or two syllables could have elicited floating stress, diffraction, or even one more "beat". Hence, we cannot distinguish which specific kind of rhythmic discrepancy led to which specific eye-movement reaction. It might be a challenging task to integrate such uniformity into these kinds of long (poetic) text stimuli, particularly because the rhythm of each traditionally composed poem is individual and is therefore describable only in comparison with a projected meter (see Burdorf, 1997, p. 69-73 for a discussion of rhythm in poetry).

Third, since each type of anomaly always occurred at the same position, effects of anomaly type might be confounded with other position-related variables, such as practice effects. This possibility needs to be addressed in future research, which is why another experimental design has been set up including a rotation system of manipulations. It will be published elsewhere. However, since the MRRL-text differed only slightly across different conditions, we believe that it might be very demanding to mentally represent and update the beat or rhythmic structure in the poem (or prose) layout while, at the same time, remembering where exactly which rhythmic deviation had occurred.

Fourth, a potential weakness of the analysis is that other lexical variables of successor words were not included. Word length and frequency, for instance, could affect parafoveal word recognition and saccade planning. However, since we were mainly interested in pronunciation-related parafoveal processing, other variables were not additionally analyzed. Also, a variable 'position within line' and 'stanza' could have been integrated to analyze where, within a stanza, reading might change, especially since the stanza structure repeats periodically. This should be considered in replications of the study.

Conclusion

We found several indicators of rhythmical subvocalization in silent MRRL reading. Metrical anomalies elicited longer single fixation durations, gaze durations, and regression path durations, while rhyme anomalies triggered re-reading of the rhyme-primes. If readers had not picked up the rhythmical structure from the stimuli, our meter and rhyme manipulations would have gone unnoticed and no difference in reading for the consistent and inconsistent versions should have been measured. The clear anomaly effects thus strongly suggest rhythmic subvocalization in silent MRRL reading.

Strict verse-by-verse poem layout can strengthen rhythmic expectations, leading to robust meter anomaly effects. MRRL in prose layout, on the other hand, elicited rhyme effects, indicating that the language was also processed rhythmically. However, due to the lack of visual formal cues, rhymes were utilized as important anchor points to establish the poetic gestalt.

With respect to general parameters indicating rhythmic subvocalization in silent reading, we argued that effects of pronounceability, such as a high correlation of fixation duration measures with the pronunciation length of words, as measured by the number of syllables, clearly speak in favor of a close alignment of eye-movements and the inner voice. Moreover, we found evidence for parafoveal processing of features related to pronounceability, further supporting the subvocalization assumption. Future research will have to show whether this finding extends to other variables related to pronunciation such as neighborhood density or sonority (Xue et al., 2019, 2020).

Our results suggest that rhythmicity in subvocalization may be graded depending on how easy or difficult it is to pick up an MRRL sound "gestalt". Thus, the alignment of the eyes with the inner voice may also be graded. Future research will have to investigate the extent to which rhythmic subvocalization is evoked in non-MRRL type poems. It is possible that an inherent beat is needed for these types of effects, but this remains an open question for future research.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html and that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

We wish to thank Dennis Erath for his support in running the experiment, and Evelyn Ferstl, Simon Kirsch, Kyla McConnell and Lisa Zacharski for fruitful discussions and comments. The article processing charge was funded by the Baden-Wuerttemberg Ministry of Science, Research and Art and the University of Freiburg in the funding programme Open Access Publishing.

References

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*(7), 1059–1068. https://doi.org/10.3758/BF03205520

Alderson-Day, B., Bernini, M., & Fernyhough, C. (2017). Uncharted features and dynamics of reading: Voices, characters, and crossing of experiences. *Consciousness and Cognition*, *49*, 98–109. https://doi.org/10.1016/j.concog.2017.01.003

Andreetta, S., Soldatkina, O., Boboeva, V., & Treves, A. (2021). *In poetry, if meter has to help memory, it takes its time* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.03.14.435310

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–Beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, *25*(9), 3077–3085. https://doi.org/10.1093/cercor/bhu103

Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 191–204. https://doi.org/10.1037/aca0000033

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89-100. https://doi.org/10.1016/j.cognition.2004.12.006

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 1–46). Springer.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beier, E. J., & Ferreira, F. (2018). The temporal prediction of stress in speech and its relation to musical beat perception. *Frontiers in Psychology*, *9*, 431. https://doi.org/10.3389/fpsyg.2018.00431

Belica, C., Kupietz, M., Lüngen, H., & Perkuhn, R. (2012). *DeReWo–Korpusbasierte Grund-/Wortformenlisten.* derewo-v-ww-bll-320000g-2012-12-31-1.0.

Berent, I. (2013). The phonological mind. *Trends in Cognitive Sciences*, *17*(7), 319–327. https://doi.org/10.1016/j.tics.2013.05.004

Blohm, S., Menninghaus, W., & Schlesewsky, M. (2017). Sentence-level effects of literary genre: Behavioral and electrophysiological evidence. *Frontiers in Psychology*, *8*, 1887. https://doi.org/10.3389/fpsyg.2017.01887

Blohm, S., Wagner, V., Schlesewsky, M., & Menninghaus, W. (2018). Sentence judgments and the grammar of poetry: Linking linguistic structure and poetic effect. *Poetics*, *69*, 41–56. https://doi.org/10.1016/j.poetic.2018.04.005

Blohm, Stefan. (2020). *Literary Psycholinguistics and the Poem* [University of Mainz]. https://publications.ub.uni-mainz.de/theses/frontdoor.php?source_opus=100003271

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing: Implicit prosody. *Language and Linguistics Compass*, *8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Breen, M. (2018). Effects of metric hierarchy and rhyme predictability on word duration in The Cat in the Hat. *Cognition*, *174*, 71–81. https://doi.org/10.1016/j.cognition.2018.01.014

Breen, M., & Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*(2), 153–170. https://doi.org/10.1016/j.jml.2010.11.001

Breen, M., & Clifton, C. (2013). Stress matters revisited: A boundary change experiment. *Quarterly Journal of Experimental Psychology*, *66*(10), 1896–1909. https://doi.org/10.1080/17470218.2013.766899

Breen, M., Fitzroy, A. B., & Oraa Ali, M. (2019). Event-related potential evidence of implicit metric structure during silent reading. *Brain Sciences*, *9*(8), 192. https://doi.org/10.3390/brainsci9080192

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 306–323. https://doi.org/10.1037/a0038689

Carminati, M. N., Stabler, J., Roberts, A. M., & Fischer, M. H. (2006). Readers' responses to sub-genre and rhyme scheme in poetry. *Poetics*, *34*(3), 204–218. https://doi.org/10.1016/j.poetic.2006.05.001

Carper, T., & Attridge, D. (2003). *Meter and meaning: An introduction to rhythm in poetry*. Routledge.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Castiglione, D. (2017). Difficult poetry processing: Reading times and the narrativity hypothesis. *Language and Literature: International Journal of Stylistics*, *26*(2), 99–121. https://doi.org/10.1177/0963947017704726

Castiglione, D. (2019). *Difficulty in Poetry: A Stylistic Model*. Springer International Publishing. https://doi.org/10.1007/978-3-319-97001-1

Chafe, W. (1988). Punctuation and the Prosody of Written Language. *Written Communication*, *5*(4), 395–426. https://doi.org/10.1177/0741088388005004001

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment To Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Eiter, B. M., & Inhoff, A. W. (2010). Visual word recognition during reading is followed by subvocal articulation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(2), 457–470. https://doi.org/10.1037/a0018278

Essens, P. J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*(1), 1–7. https://doi.org/10.3758/BF03207132

Fabb, N. (2009). Why Is Verse Poetry? *PN Review*, *36*(1), 52. MLA International Bibliography.

Fabb, N. (2013). There is no psychological limit on the duration of metrical lines in performance: Against Turner and Pöppel. *International Journal of Literary Linguistics*, *2*(1). https://doi.org/10.15462/IJLL.V2I1.14

Fabb, N. (2015). *What is Poetry?: Language and Memory in the Poems of the World*. Cambridge University Press.

Fabb, N., Halle, M., & Piera, C. (2008). *Meter in poetry: A new theory*. Cambridge University Press.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491–1506. https://doi.org/10.1037/a0036858

Fechino, M., Jacobs, A. M., & Lüdtke, J. (2020). Following in Jakobson and Lévi-Strauss' footsteps: A neurocognitive poetics investigation of eye movements during the reading of Baudelaire's 'Les Chats.' *Journal of Eye Movement Research*, *13*(3). https://doi.org/10.16910/jemr.13.3.4

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9.

Fernholz, I., Menzel, J., Jabusch, H.-C., Gembris, H., Fischer, F., Kendel, F., Kreutz, G., Schmidt, A., Willich, S. N., & Weikert, C. (2018). Musikalische Inaktivität – ein Risikofaktor? Vorstellung eines kurzen Fragebogens zur Erfassung der musikalischen Aktivität (MusA). *Das Gesundheitswesen*. https://doi.org/10.1055/s-0044-101143

Fitzroy, A. B., & Breen, M. (2020). Metric structure and rhyme predictability modulate speech intensity during child-directed and read-alone productions of children's literature. *Language and Speech*, *63*(2), 292–305. https://doi.org/10.1177/0023830919843158

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulletin & Review*, *18*(4), 736–741. https://doi.org/10.3758/s13423-011-0105-x

Floessner, T., & Hut, R. A. (2017). Basic principles underlying biological oscillations and their entrainment. In V. Kumar (Ed.), *Biological Timekeeping: Clocks, Rhythms and Behaviour* (Vol. 366, pp. 47–58). Springer India.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319.

Fodor, J. D. (2002a). *Prosodic disambiguation in silent reading: M. Hirotani (ed.), Proceedings of the North East Linguistic Society 32*. GSLA, University of Massachusetts.

Fodor, J. D. (2002b). Psycholinguistics cannot escape prosody. *SPEECH PROSODY 2002 Conference*. Speech Prosody, Aix-en-Provence, France.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic Beta oscillations. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(5), 1791–1802. https://doi.org/10.1523/JNEUROSCI.4107-11.2012

Fuller, D. (2001). *Pause*. Oxford University Press. https://doi.org/10.1093/gmo/9781561592630.article.21118

Gjerdingen, R. O. (1989). Meter as a mode of attending: A network simulation of attentional rhythmicity in music. *Intégral: The Journal of Applied Musical Thought*, *3*, 67–91. RILM Abstracts of Music Literature (1967 to present).

Grimm, J. und W. (1884). *Household tales by the Brothers Grimm. Jacob and Wilhelm Grimm* (M. Hunt, Trans.). George Bell and Sons.

Hanauer, D. (1996). Integration of phonetic and graphic features in poetic text categorization judgements. *Poetics*, *23*(5), 363–380. https://doi.org/10.1016/0304-422X(95)00010-H

Hanauer, D. (1998a). Reading Poetry: An Empirical Investigation of Formalist, Stylistic, and Conventionalist Claims. *Poetics Today*, *19*(4), 565. https://doi.org/10.2307/1773260

Hanauer, D. (1998b). The genre-specific hypothesis of reading: Reading poetry and encyclopedic items. *Poetics*, *26*(2), 63–80. https://doi.org/10.1016/S0304-422X(98)00011-4

Hanauer, D. (2001). What we know about reading poetry. In D. Schram & G. J. Steen (Eds.), *The Psychology and Sociology of Literature* (Vol. 35, pp. 107–128). John Benjamins Publishing Company. https://doi.org/10.1075/upal.35.08han

Hetherington, P., & Atherton, C. L. (2020). *Prose poetry: An introduction*. Princeton University Press.

Hoffstaedter, P. (1987). Poetic text processing and its empirical investigation. *Poetics*, *16*(1), 75–91. https://doi.org/10.1016/0304-422X(87)90037-4

Honing, H. (2012). Without it no music: Beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*, 85–91. https://doi.org/10.1111/j.1749-6632.2011.06402.x

Honing, H. (Ed.). (2018). *The origins of musicality*. The MIT Press.

Honing, H. (2019). *Der Affe schlägt den Takt*. Henschel Verlag.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Huey, E. (1908). *The psychology and pedagogy of reading*. Macmillan. https://pure.mpg.de/rest/items/item_2347473_2/component/file_2347472/content

Hurschler, M. A., Liem, F., Oechslin, M., Stämpfli, P., & Meyer, M. (2015). FMRI reveals lateralized pattern of brain activity modulated by the metrics of stimuli during auditory rhyme processing. *Brain and Language*, *147*, 41–50. https://doi.org/10.1016/j.bandl.2015.05.004

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kayser, W. (2002). *Kleine deutsche Versschule*. UTB GmbH.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, *123*(1), 1–20. https://doi.org/10.1016/j.cognition.2011.11.012

Kentner, G. (2016). Rhythmic parsing. *The Linguistic Review*, *34*(1), 123–155. https://doi.org/10.1515/tlr-2016-0005

Kentner, G., & Vasishth, S. (2016). Prosodic focus marking in silent reading: Effects of discourse context and rhythm. *Frontiers in Psychology*, *7*, 319. https://doi.org/10.3389/fpsyg.2016.00319

Kien, J. & Kemp., A. (1994). Is speech temporally segemented? Comparison with temporal segmentation in behaviour. *Brain and Language*, *46*, 662–682.

Kiparsky, P. (2009). Review: Meter in Poetry: A New Theory by Nigel Fabb and Morris Halle. *Language*, *85*(4), 923–930.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Knoop, C. A., Blohm, S., Kraxenberger, M., & Menninghaus, W. (2019). How perfect are imperfect rhymes? Effects of phonological similarity and verse context on rhyme perception. *Psychology of Aesthetics, Creativity, and the Arts.* https://doi.org/10.1037/aca0000277

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Konieczny, L., Hemforth, B., & Scheepers, C. (2000). Head position and clause boundary effects in reanalysis. In B. Hemforth & L. Konieczny (Eds.), *German Sentence Processing* (pp. 247–278). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-9618-3_

Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes*, *12*, 307–348. https://doi.org/10.1080/016909697386871

Koops van ’t Jagt, R., Hoeks, J., Dorleijn, G. J., & Hendriks, P. (2014). Look before you leap: How enjambment affects the processing of poetry. *Scientific Study of Literature*, *4*(1), 3–24. https://doi.org/10.1075/ssol.4.1.01jag

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*(18), 2867–2875. https://doi.org/10.1016/j.cub.2018.07.023

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kotz, S. A., & Schwartze, M. (2016). Motor-timing and sequencing in speech production: A general-purpose framework. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 717–724). Academic Press.

Kraxenberger, M., Menninghaus, W., Roth, A., & Scharinger, M. (2018). Prosody-based sound-emotion associations in poetry. *Frontiers in Psychology*, *9*, 1284. https://doi.org/10.3389/fpsyg.2018.01284

Kriukova, O., & Mani, N. (2016). The strong, the weak, and the first: The impact of phonological stress on processing of orthographic errors in silent reading. *Brain Research*, *1636*, 208–218. https://doi.org/10.1016/j.brainres.2016.01.003

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PloS ONE*, *10*(11), e0141069. https://doi.org/10.1371/journal.pone.0141069

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lange, E. B., Pieczykolan, A., Trukenbrod, H. A., & Huestegge, L. (2018). *The rhythm of cognition – Effects of an auditory beat on oculomotor control in reading and sequential scanning*. *11*(2). https://doi.org/10.16910/jemr.11.2.9

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience and Biobehavioral Reviews*, *81*(Pt B), 158–166. https://doi.org/10.1016/j.neubiorev.2016.12.012

Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01432

Lee, H.-W., Rayner, K., & Pollatsek, A. (2002). The processing of consonants and vowels in reading: Evidence from the fast priming paradigm. *Psychonomic Bulletin & Review*, *9*(4), 766–772. https://doi.org/10.3758/BF03196333

Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences*, *930*(1), 337–354. https://doi.org/10.1111/j.1749-6632.2001.tb05743.x

Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In *Arbib, M.A. (ed.): Language, Music, and the Brain: A Mysterious Relationship* (pp. 257–272). MIT Press.

Lerdahl, F., & Jackendoff, R. (1983). An Overview of Hierarchical Structure in Music. *Music Perception: An Interdisciplinary Journal*, *1*(2), 229–252. https://doi.org/10.2307/40285257

London, J. (2012a). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). Oxford University Press.

London, J. (2012b). Three things linguists need to know about rhythm and time in music. *Empirical Musicology Review*, *7*(1–2), 5–11. https://doi.org/10.18061/1811/52973

Manguel, A. (1996). *A history of reading*. Flamingo.

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Menninghaus, W., & Blohm, S. (2020). Empirical aesthetics of poetry. In M. Nadal & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.33

Menninghaus, W., Wagner, V., Knoop, C. A., & Scharinger, M. (2018). Poetic speech melody: A crucial link between music and language. *PloS ONE*, *13*(11). https://doi.org/10.1371/journal.pone.0205980

Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics*, *63*, 47–59. https://doi.org/10.1016/j.poetic.2016.12.001

Menninghaus, W., & Wallot, S. (2021). What the eyes reveal about (reading) poetry. *Poetics*, 101526. https://doi.org/10.1016/j.poetic.2020.101526

Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*(1), 4–17. https://doi.org/10.1016/j.cortex.2008.06.011

Metz-Göckel, H. (Ed.). (2008). *Gestalttheorie aktuell* (1. ed., Vol. 1). Krammer.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23–34. https://doi.org/10.1016/j.cognition.2018.12.015

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1658), 20130396. https://doi.org/10.1098/rstb.2013.0396

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1658), 20130393. https://doi.org/10.1098/rstb.2013.0393

Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective & Behavioral Neuroscience*, *16*(2), 362–373. https://doi.org/10.3758/s13415-015-0396-x

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, *4*, 1–10. https://doi.org/10.3389/fpsyg.2013.00010

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Patel, A. D. (2003). Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences*, *999*, 140–143. https://doi.org/10.1196/annals.1284.015

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Løevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Peskin, J. (2007). The genre of poetry: Secondary school students' conventional expectations and interpretive operations. *English in Education*, *41*(3), 20–36. https://doi.org/10.1111/j.1754-8845.2007.tb01162.x

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*(3–4), 599–611. https://doi.org/10.1016/j.wocn.2003.08.001

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.com

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. A. (2019). Rhythm in speech and animal vocalizations: A cross-species perspective. *Annals of the New York Academy of Sciences*, *1453*(1), 79–98. https://doi.org/10.1111/nyas.14166

Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, *8*, 1820. https://doi.org/10.3389/fpsyg.2017.01820

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice-Hall.

Rebuschat, P., Rohmeier, M., Hawkins, J. A., & Cross, I. (Eds.). (2011). *Language and Music as Cognitive Systems*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553426.001.0001

Ringelnatz, J. (1997). *Sämtliche Gedichte*. Diogenes.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Rubin, D. C., Ciobanu, V., & Langston, W. (1997). Children's memory for counting-out rhymes: A cross-language comparison. *Psychonomic Bulletin & Review*, *4*(3), 421–424. https://doi.org/10.3758/BF03210804

Scheepers, C., Mohr, S., Fischer, M. H., & Roberts, A. M. (2013). Listening to Limericks: A pupillometry investigation of perceivers' expectancy. *PloS ONE*, *8*(9), e74986. https://doi.org/10.1371/journal.pone.0074986

Schmidt-Kassow, M., & Kotz, S. A. (2009a). Attention and perceptual regularity in speech. *Neuroreport*, *20*(18), 1643–1647. https://doi.org/10.1097/WNR.0b013e328333b0c6

Schmidt-Kassow, M., & Kotz, S. A. (2009b). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21*(9), 1693–1708. https://doi.org/10.1162/jocn.2008.21153

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, *5*(80). https://doi.org/10.3389/fpsyg.2014.00080

Schofield, M. (Ed.). (2016). *Laws: Plato*. Cambridge University Press.

Schrott, R., & Jacobs, A. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren*. Carl Hanser Verlag.

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in context: The effects of length, frequency, and predictability on brain responses during natural reading. *Cerebral Cortex*, *26*(10), 3889.2-3904. https://doi.org/10.1093/cercor/bhw184

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, *3*, 371–405. https://doi.org/10.1017/S0952675700000695

Silva, S., & Castro, S. L. (2019). The time will come: Evidence for an eye-audiation span in silent music reading. *Psychology of Music*, *47*(4), 504–520. https://doi.org/10.1177/0305735618765302

Skov, M., & Nadal, M. (2020). There are no aesthetic emotions: Comment on Menninghaus et al. (2019). *Psychological Review*, *127*(4), 640–649. https://doi.org/10.1037/rev0000187

Slana, A., Repovš, G., Fitch, W. T., & Gingras, B. (2016). Harmonic context influences pitch class equivalence judgments through gestalt and congruency effects. *Acta Psychologica*, *166*, 54–63. https://doi.org/10.1016/j.actpsy.2016.03.006

Smith, B. H. (1968). *Poetic closure: A study of how poems end*. University of Chicago Press.

Stolterfoht, B., Friederici, A., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition*, *104*(3), 565–590. https://doi.org/10.1016/j.cognition.2006.08.001

Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(38), 14981–14988. https://doi.org/10.1523/JNEUROSCI.0612-13.2013

Tierney, A., & Kraus, N. (2015). Neural entrainment to the rhythmic structure of music. *Journal of Cognitive Neuroscience*, *27*(2), 400–408. https://doi.org/10.1162/jocn_ a_ 00704

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience*, *29*(5), 855–868. https://doi.org/10.1162/jocn_a_01092

Tillmann, B., & Jay Dowling, W. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition*, *35*(4), 628–639. https://doi.org/10.3758/BF03193301

Tilsen, S. (2019). Space and time in models of speech rhythm. *Annals of the New York Academy of Sciences*, *1453*(1), 47–66. https://doi.org/10.1111/nyas.14102

Tremblay, A. & Ransijn, J. (2015). *LMERConvenience-Functions: Model selection and post-hoc analysis for (G)LMER models (Version 2.10).* https://CRAN.R-project.org/package = LMER ConvinienceFunctions

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Tsur, R., Glicksohn, J., & Goodblatt, C. (1991). Gestalt qualities in poetry and the reader's absorption style. *Journal of Pragmatics*, *16*(5), 487–500. https://doi.org/10.1016/0378-2166(91)90138-N

Turner, F., & Pöppel, E. (1983). The neural lyre: Poetic meter, the brain and time. *Poetry*, *142*(5), 277–309. https://www.jstor.org/stable/20599567

Van Peer, W. (1990). The measurement of metre: Its cognitive and affective functions. *Poetics*, *19*(3), 259–275.

Vilhauer, R. P. (2017). Characteristics of inner reading voices. *Scandinavian Journal of Psychology*, *58*(4), 269–274. https://doi.org/10.1111/sjop.12368

Wade, B. C. (2004). *Thinking musically: Experiencing music, expressing culture*. Oxford University Press.

Wang, L., Bao, Y., Zhang, J., Lin, X., Yang, L., Pöppel, E., & Zhou, B. (2016). Scanning the world in three seconds: Mismatch negativity as an indicator of temporal segmentation: Mismatch negativity and temporal segmentation. *PsyCh Journal*, *5*(3), 170–176. https://doi.org/10.1002/pchj.144

Wang, L., Lin, X., Zhou, B., Pöppel, E., & Bao, Y. (2015). Subjective present: A window of temporal integration indexed by mismatch negativity. *Cognitive Processing*, *16*, 131–135. https://doi.org/10.1007/s10339-015-0687-8

Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, *12*(8), 1229–1240. https://doi.org/10.1093/scan/nsx069

Whittall, A. (2011). Rhythm. In A. Latham (Ed.), *The Oxford Companion to Music*. Oxford University Press.

Xue, S., Jacobs, A. M., & Lüdtke, J. (2020). What Is the Difference? Rereading Shakespeare's Sonnets—An Eye Tracking Study. *Frontiers in Psychology*, *11*, 49. https://doi.org/10.3389/fpsyg.2020.00421

Xue, S., Lüdtke, J., Sylvester, T., & Jacobs, A. M. (2019). Reading Shakespeare sonnets: Combining quantitative narrative analysis and predictive modeling - an eye tracking study. *Journal of Eye Movement Research*, *12*(5). https://doi.org/10.16910/jemr.12.5.2

Yaron, I. (2002). Processing of obscure poetic texts: Mechanisms of selection. *Journal of Literary Semantics*, *31*(2), 178. https://doi.org/10.1515/jlse.2002.013

Yaron, I. (2008). What is a "difficult poem"? Towards a definition. *Journal of Literary Semantics*, *37*(2), 107. https://doi.org/10.1515/jlse.2008.008

Yoncheva, Y. N., Maurer, U., Zevin, J. D., & McCandliss, B. D. (2013). Effects of rhyme and spelling patterns on auditory word ERPs depend on selective attention to phonology. *Brain and Language*, *124*(3), 238–243. https://doi.org/10.1016/j.bandl.2012.12.013

Yu, X., & Bao, Y. (2020). The three second time window in poems and language processing in general: Complementarity of discrete timing and temporal continuity. *PsyCh Journal*, *9*(4), 429–443. https://doi.org/10.1002/pchj.390

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Current Biology*, *28*(3), 401-408.e5. https://doi.org/10.1016/j.cub.2017.11.071

Zwaan, R. A. (1991). Some parameters of literary and news comprehension: Effects of discourse-type perspective on reading rate and surface structure representation. *Poetics*, *20*(2), 139–156. https://doi.org/10.1016/0304-422X(91)90003-8