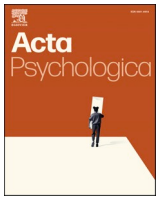
Acta Psychologica 235 (2023) 103892

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



Distinguishing between struggling and skilled readers based on their prosodic speech patterns in oral reading: An exploratory study in grades 2 and 4☆

Panagiotis Karageorgos [a,*], Sebastian Wallot [b], Bettina Müller [a,c], Julia Schindler [a], Tobias Richter [a]

[a] Department of Psychology IV, University of Würzburg, Germany
[b] Department of Methodology and Evaluation Research, Leuphana University Lüneburg, Germany
[c] Competence Center for School Psychology Hesse, Goethe University Frankfurt am Main, Germany

ARTICLE INFO

Keywords:
Recurrence quantification analysis
Reading prosody
German primary school

ABSTRACT

The purpose of this study was to examine if prosodic patterns in oral reading derived from Recurrence Quantification Analysis (RQA) could distinguish between struggling and skilled German readers in Grades 2 ($n = 67$) and 4 ($n = 69$). Furthermore, we investigated whether models estimated with RQA measures outperformed models estimated with prosodic features derived from prosodic transcription. According to the findings, struggling second graders appear to have a slower reading rate, longer intervals between pauses, and more repetitions of recurrent amplitudes and pauses, whereas struggling fourth graders appear to have less stable pause patterns over time, more pitch repetitions, more similar amplitude patterns over time, and more repetitions of pauses. Additionally, the models with prosodic patterns outperformed models with prosodic features. These findings suggest that the RQA approach provides additional information about prosody that complements an established approach.

1. Introduction

Thinking back to our school days, we all remember classmates who, in comparison to their peers, struggled while reading a passage aloud. These students read passages word for word, syllable by syllable, or even letter by letter, with long pauses in between, and they took longer than most of their peers to complete reading the passage. This observable behavior has been described as a lack of reading fluency in reading research and by literacy educators (Allington, 1983; Dowhower, 1991).

Reading fluency has received considerable attention since the Report of the National Institute of Child Health and Human Development (National Institute of Child Health and Human Development, 2000). However, providing a widely accepted definition appears rather challenging (Godde et al., 2020). A consensus exists that reading fluency is a complex, multifaceted construct, which mainly relies on word recognition accuracy, automaticity, and prosody (e.g., Hudson et al., 2009; Kuhn et al., 2010). Nonetheless, the contribution of each factor to reading fluency remains disputed (Kuhn et al., 2010). Despite the ample research on reading fluency, it tends to focus on accuracy and automaticity (e.g., Fuchs et al., 2001; Kara et al., 2020), whereas prosody is frequently overlooked (Silverman et al., 2013; Wolters et al., 2020). According to Dowhower (1991), one possible explanation for the neglect of prosody could be that measuring word reading accuracy and word reading speed is easier and less time consuming than measuring prosody. In the present study, we employed *Recurrence Quantification Analysis* (RQA – Marwan et al., 2007; Wallot et al., 2014) and *k-fold cross-validation analysis* (Browne, 2000) to explore the differences in the prosodic patterns of skilled and struggling German readers in second and fourth grade. We were particularly interested in identifying prosodic patterns that can distinguish between skilled and struggling German readers. Furthermore, we investigated whether these prosodic patterns could distinguish between struggling and skilled German readers better

☆ This article is based on data published in Müller et al. (2015). We have no known conflict of interest to disclose. The research reported in this article was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under Grant 01GJ1004. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

* Corresponding author at: University of Würzburg, Department of Psychology IV, Röntgenring 10, Würzburg 97070, Germany.
E-mail address: panagiotis.karageorgos@uni-wuerzburg.de (P. Karageorgos).

https://doi.org/10.1016/j.actpsy.2023.103892
Received 5 October 2022; Received in revised form 6 February 2023; Accepted 13 March 2023
Available online 24 March 2023

than prosodic features, such as appropriate and inappropriate pause occurrences acquired from prosodic transcription. In the following sections, we briefly describe word-reading automaticity, word-reading prosody, and the presently available methods to assess prosody, which provides the groundwork for our study.

1.1. Word Reading automaticity and prosody

Many researchers regard fluency as more or less synonymous with word reading automaticity (Dowhower, 1991). Word reading automaticity reflects the ability to recognize words accurately and with minimal cognitive effort (Paige et al., 2014). The number of words correctly read per minute, which incorporates both word reading accuracy and word reading speed, is a common metric of word reading automaticity (Benjamin & Schwanenflugel, 2010). Nevertheless, achieving a word recognition accuracy threshold is an important precondition before word reading speed starts improving (Altani et al., 2020; Juul et al., 2014; Karageorgos et al., 2019, 2020). For readers who lack a high level of word reading accuracy, word recognition will most likely be inefficient, resulting in a high demand on cognitive resources and slow reading speed (Castles et al., 2018; LaBerge & Samuels, 1974). Hence, before readers recognize words efficiently, they must first learn the grapheme-phoneme correspondence rules and start making use of syllables (Coltheart et al., 2001).

Even though word reading automaticity is necessary for reading fluency and reading comprehension in early primary school (e.g., Ehri, 2005; Miller & Schwanenflugel, 2008), it is not sufficient (Kuhn et al., 2010). Prosody, which "refers to reading with expression" (Rasinski et al., 2009, p. 351), is also considered by many reading researchers to be a key component of reading fluency (e.g., Dowhower, 1991; Hudson et al., 2009). Several studies have found a direct positive relationship between prosody and reading comprehension (e.g., Benjamin & Schwanenflugel, 2010; Rasinski et al., 2009). As a result, reading prosody may deserve closer attention than it currently receives.

According to Kuhn et al. (2010; see also Couper-Kuhlen, 1986), four well-established prosodic features describe the quality of prosody as variations in a) *pitch*, b) *duration*, c) *stress*, and d) *pausing*. *Pitch*, which is also referred to as intonation or fundamental frequency, could be described as the rate of vibration of a speaker's vocal folds (Dowhower, 1991). Pitch appears to depend on a variety of factors such as the language or even the dialect of the same language (e.g., Mennen et al., 2014). In a cross-language comparison, Mennen et al. (2012) found significant differences in pitch span between English and German female monolingual speakers. English female speakers had a higher pitch at the start of a phrase and showed more pitch variations than German female speakers (see also Scharff-Rethfeldt et al., 2008). Given that no apparent organic or physiological differences existed between the two groups, pitch changes could be traced back to the spoken language (Mennen et al., 2012, 2014).

Pitch may also vary depending on a speaker's gender, with men having lower fundamental frequencies than women (e.g., Hillenbrand & Clark, 2009; Lee et al., 1999; Titze, 1989). Lee et al. (1999) found in their study with children speaking American English that these differences begin to manifest around the age of 11 and are fully established by the age of 15. Finally, comparisons between skilled and struggling readers in primary school have shown that skilled readers have a larger pitch range and more appropriate pitch falls and rises than struggling readers (e.g., English: Benjamin et al., 2013; Spanish: Álvarez-Cañizo et al., 2015).

The second prosodic feature, *duration,* refers to the pronunciation length of vowels and syllables and is usually measured in milliseconds (Kuhn et al., 2010). Duration may depend on a variety of factors. For example, vowels in stressed words have longer pronunciations than vowels in unstressed words (Temperley, 2009). A vowel's pronunciation can even be longer during the phrase-final lengthening, which is the lengthening of the vowel at the final position of a phrase (e.g., Dowhower, 1987, 1991; Edwards et al., 1991; Turk & Shattuck-Hufnagel, 2007). Research has shown that appropriate phrase-final lengthening is a good indicator of readers chunking the reading material (Cooper & Cooper, 1980; Dowhower, 1991), which enhances reading comprehension (Stevens, 1981). Nevertheless, when comparing pronunciations between readers, their individual speaking rate should also be assessed. Faster readers have a faster speaking rate and therefore shorter segment durations than slower readers (Kuhn et al., 2010).

The third prosodic feature, *stress,* refers to the phonetical accentuation with which a vowel or a word is read and is usually measured in dB (Ktori et al., 2018). Stressing a syllable is particularly important because it aids in the distinction of words at the word and sentence level (Ktori et al., 2018). For example, stressing the first or second syllable in English could indicate whether the word is a noun or a verb (e.g., the noun *per*mit versus the verb per*mit*; Himmelmann & Ladd, 2008). Rhythmic patterns and rules may differ between languages (Kuhn et al., 2010). In some languages, such as Greek and French, special diacritics on written words denote the syllables that should be stressed when reading a polysyllabic word (Protopapas, 2006), but most written systems lack such diacritics and stress might be determined by other rules such as morphemes (Rastle & Coltheart, 2000; see also Ktori et al., 2018). Finally, comparisons between English skilled and struggling readers in primary school showed differences in the frequency of stresses. Skilled readers insert a stress every 4.7 words, whereas struggling readers insert a stress in almost every word (Clay & Imlach, 1971; Dowhower, 1991).

The last important prosodic feature, *pausing*, refers to the intra-sentential and inter-sentential pauses that occur during reading and is indicated by a silence in the signal (Kuhn et al., 2010). According to Lalain et al. (2016, as cited in Godde et al., 2020), pauses can be distinguished into three types: breath pauses, syntactic pauses, and hesitation pauses. Breath pauses are indispensable for air intake and may also be used as discourse markers (Bailly & Gouvernayre, 2012), whereas syntactic pauses support sentence parsing and reading comprehension, and hesitation pauses indicate cognitive activity and are associated with decoding (Godde et al., 2020). Most studies focus on the duration, frequency, and appropriateness of the pauses (e.g., English: Schwanenflugel et al., 2004; French: Godde et al., 2022), which are determined to a large extent by the grammatical structure of the text (Bailly & Gouvernayre, 2012; Miller & Schwanenflugel, 2008). In addition, the duration of pauses may also vary depending on the spoken language. For example, Italian has a shorter average pause duration than English, German, and French, whereas Spanish has a longer average pause duration than these three languages (Campione & Véronis, 2002). Several findings with primary school children also suggest that skilled readers take fewer and shorter pauses (e.g., Miller & Schwanenflugel, 2006; Schwanenflugel et al., 2004), whereas less skilled readers make more inappropriate pauses (e.g., Álvarez-Cañizo et al., 2015; Miller & Schwanenflugel, 2008).

In sum, despite differences between the languages in the contribution of the prosodic features to reading, some similarities can be observed when comparing skilled readers to struggling readers. Skilled readers in primary school show larger pitch ranges and shorter pause durations, more appropriate pitch variations, lengthening and stresses, and fewer pauses and stresses compared to struggling readers.

1.2. Present methods for assessing prosody

Reading prosody rating scales, spectrographic analyses, and prosodic transcription are presently the most common methods used for measuring and analysing prosodic qualities (e.g., Kuhn et al., 2010; Wolters et al., 2020). However, as it is the case with every research method, they come with advantages and disadvantages. In reading prosody rating scales, human raters make subjective judgements on how readers perform according to certain criteria rubrics (Morrison & Wilcox, 2020). Several reading prosody rating scales exist, each one focusing on different aspects (e.g., Benjamin et al., 2013; Pinnell et al.,

1995; Rasinski, 2004). According to Wolters et al. (2020), the most widely used reading prosody scales are the NAEP oral reading fluency scale (Daane et al., 2005; Pinnell et al., 1995) and the Multidimensional Fluency Scale (Rasinski, 2004). The NAEP is a 4-point holistic rubric that measures phrasing, adherence to the author's syntax, and expressiveness (Daane et al., 2005; Pinnell et al., 1995). The Multidimensional Fluency Scale is a 16-point analytic rubric, with four dimensions and four criteria per dimension. Readers are rated in the dimensions of expression and volume, phrasing, smoothness, and pace. The main advantage of these rating scales is that they are convenient tools for teachers and can be used in the classroom (Miller & Schwanenflugel, 2006). However, some of these scales require the raters to be trained before they can use them, and rating a single reading session can be time consuming.

In spectrographic analysis, sound waves are transformed into a visual representation which is known as a spectrogram (Schwanenflugel et al., 2004). Many features, such as stress, pitch, and pause lengths, can be extracted using this representation (Molholt, 1990; Schwanenflugel et al., 2004). Advances in technology and new software have made spectrographic analysis easier, and it allows for a more precise assessment of prosodic features (e.g., Praat by Boersma & Weenink, 2021), but it is more complex to use than reading rating scales. The required technical skills often exceed those of teachers and reading specialists (Kuhn et al., 2010). Furthermore, even though spectrograms are a powerful tool, researchers must still examine spectrograms to discern patterns in the signal (Molholt, 1990).

In prosodic transcriptions, transcribers identify and annotate all prosodic features by employing a transcription system while listening to an audio file. There are several well-known transcription systems for prosody, for example, Tones and Break Indices (ToBi, Silverman et al., 1992) for American English, the Gesprächsanalytisches Transkriptionssystem (conversation-analytic transcription system, GAT, Selting et al., 1998) for German, and the International Transcription System for Intonation (INTSINT, Hirst, 1991) for cross-linguistic comparisons (Bressem, 2013). These transcription systems divide utterances into prosodically marked segments and represent prosodic aspects such as changes in pitch, accentuation, and lengthening (Bressem, 2013). However, prosodic annotation can be rather challenging. Transcription of a signal not only requires much time, it also requires the transcribers to be familiar with the corresponding transcription system to avoid errors.

1.3. The present study

The primary aim of the present study was to investigate potential differences in prosodic features and prosodic patterns between skilled and struggling German readers in primary school by applying the nonlinear analysis tool RQA (see Wallot, 2017, for an extensive tutorial). Although conventional methods, such as prosodic transcription and spectrograms, allow for the investigation of this research question, these methods have limitations. The advantage of RQA over these methods is the possibility to quantify the complexity of a time-series such as the extent that a signal is repetitive, noisy, or stationary. This quantitative method produces various metrics that accurately and objectively represent a signal's patterns (Wallot, 2017) without the need to manually search for them in a spectrogram. Thus, with this robust semi-automated process, identifying patterns in a signal is faster and less error prone. We expected that signals of skilled German readers would be more structured and stable and would show a less complex and chaotic pattern in all prosodic measures with fewer and shorter pauses and stresses than the signals of struggling German readers, which is analogous to findings from reading research in which the more structured word reading times and eye movements are associated with higher reading skills and comprehension (Tschense & Wallot, 2022; Wallot et al., 2014).

A secondary aim was to investigate whether the metrics obtained with RQA, such as signal patterns and complexity, could better distinguish between skilled and struggling readers than metrics derived from prosodic transcription. By prosodic transcription, transcribers must listen to each signal carefully to be able to identify and annotate each prosodic feature. Thus, the method yields data about the frequency of prosodic feature occurrences but no information on prosodic patterns. In contrast, RQA identifies and quantifies prosodic patterns that cannot be easily perceived with the human eye, but whether prosodic patterns can distinguish between skilled and struggling readers better than the prosodic features obtained with prosodic transcription is still unclear. RQA is strongly data-driven and allows for a semi-automated pattern detection, whereas prosodic transcription is more theory-driven and requires manual coding of each piece of information. Thus, the importance of RQA metrics regarding the classification to skilled or struggling reader compared to prosodic features obtained with prosodic transcription should also be investigated.

2. Method

2.1. Participants, procedure and design

Oral reading data were collected in a subsample of a longitudinal study that investigated differential effects of three types of reading trainings in peer-tutored learning settings (Müller et al., 2015). A subsample of 67 children in Grade 2 from seven schools (32 girls and 35 boys, aged from 7.35 to 9.19 years, $M = 8.04$ years, $SD = 0.34$) and 69 children in Grade 4 from eight schools (38 girls and 31 boys, aged from 9.20 to 12.02 years, $M = 10.14$ years, $SD = 0.57$) were selected to participate in the present study. Half of the children per grade received a reading fluency training, the other half was in the control condition and received a training of visuospatial working memory. All parents gave a written informed consent for the participation of their children. For Grade 2, parents of 49 children reported that their children's first language was German, parents of 12 children reported that the first language of their children was another language than German and parents of six children did not share any information about the first language. For Grade 4, parents of 43 children reported that their children's first language was German, parents of 14 children reported that the first language of their children is other than German and parents of 12 children did not share any information regarding the first language.

The study was based on an experimental pre−/posttest design with randomization at the class level and was conducted in the urban area of Kassel, Germany. Participants were first screened with ELFE 1–6 (Lenhard & Schneider, 2006) for reading comprehension (see Fig. 1). Five children with the lowest scores (hereafter referred to as tutees) and five children with the highest scores (hereafter referred to as tutors) in each class were chosen for the training. To hold skill differences between tutees and tutors constant, the allocation of the children in pairs was based on their score ranking. The highest-scoring tutee was paired with the highest-scoring tutor, the second-best tutee with the second-best tutor, and so forth. In Grade 2, this procedure resulted in 33 tutees and 34 tutors, and in Grade 4, it resulted in 35 tutees and 34 tutors. The proportion of girls and boys in the tutees and tutors groups was approximately equal for Grade 2, $\chi^2(1) = 0.02$, $p = .90$, and Grade 4, $\chi^2(1) = 0.74$, $p = .39$.

2.2. Measures

2.2.1. Reading comprehension

Reading comprehension skill was assessed with the computer version of the subtest text comprehension of ELFE 1–6 (Lenhard & Schneider, 2006). The test is a standardized reading comprehension test and is widely used in Germany. It consists of 13 short narrative and expository texts (two to five sentences) with one to three single-choice questions each consisting of four items. The test score is the sum of correct responses achieved within 7 min.

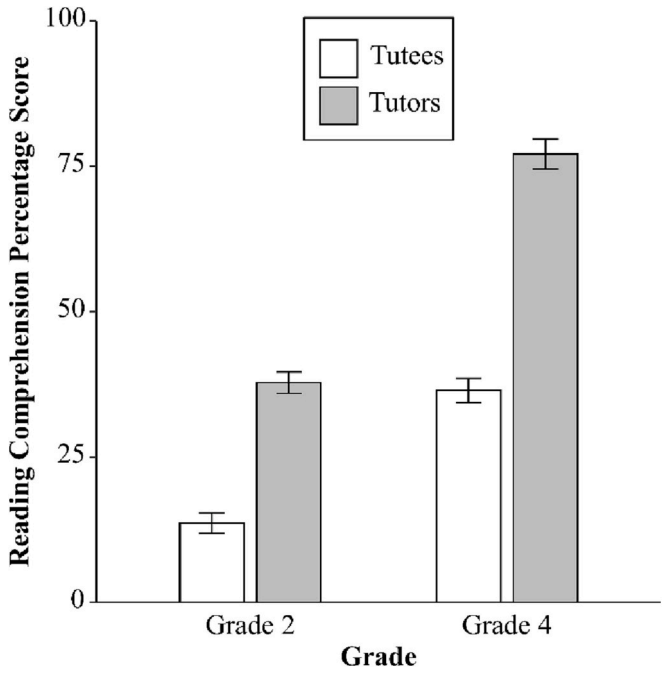


Fig. 1. Average reading comprehension percentage scores of struggling readers (tutees) and skilled readers (tutors) are shown for each grade. Reading comprehension percentage scores of the sum of correct answers in ELFE 1–6 (Lenhard & Schneider, 2006). Error bars show standard errors.

2.2.2. Prosody measures

2.2.2.1. Data collection. The training consisted of 25 sessions, each lasting 45 min and occurring in addition to regular school classes twice a week. At the 23rd training session children read a story aloud individually while they were being recorded (this issue will be addressed in the discussion). At the end of the session, children were required to answer two open-ended questions on the story they had just read (see Fig. 2). The story for the second graders was 66 words long (9 sentences) and the story for the fourth graders was 108 words long (12 sentences). The LIX readability measure of the passage for the second graders was 20.97, and of the passage for the fourth graders was 22.89. LIX is a measure of text difficulty (Bamberger & Rabin, 1984; Björnsson, 1968). The measure is computed by summing the average length of the sentences in a text and the percentage of long words with more than six letters. In German, a score of less than 25 indicates that the text has a low complexity and is appropriate for first graders (Bamberger & Rabin, 1984)

Each school had a quiet room where the children read the story, with only the student assistant and the child present. The story was written on a DIN A4 sheet of paper with the font style Calibri, font size 18, and 1.15 line spacing. During the session, children wore a headset that was connected to a laptop. The audio was digitally recorded using the Audacity software (Version 1.2.6; Audacity Team, 2006).

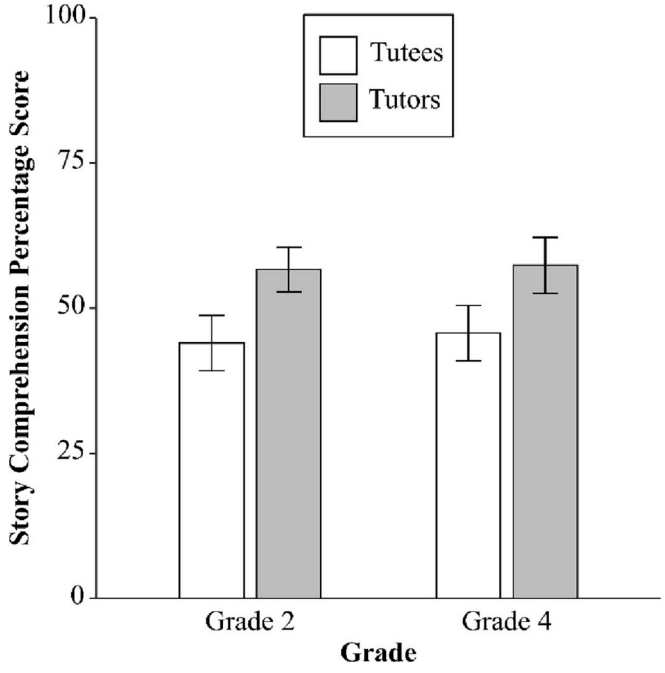


Fig. 2. Average story comprehension percentage scores of struggling readers (tutees) and skilled readers (tutors) are shown for each grade. Story comprehension percentage scores of the sum of correct answers in the story open-ended questions. Error bars show standard errors.

2.2.2.2. Data processing for RQA. All signals were processed via a custom script in the program MATLAB (Version R2017a; The Math Works, 2017). First, in an attempt to reduce the noise in each signal, a linear smoothing with equal weight in windows of a 50th of 44,100 Hz was conducted. Afterwards, boundaries of pronunciation onsets and offsets using an amplitude criterion (see MATLAB script in online supplement) were detected and a cut-off criterion of 50 ms was used to identify and remove all pronunciation periods with a shorter duration. Hence, sounds such as lip smacking were removed from the signal. Finally, all remaining pronunciation periods were concatenated to have a continuous signal. This signal was partitioned into the pauses and the pronunciation periods to investigate them separately. Pitch was extracted with an algorithm based on the Voice Analysis Toolbox in MATLAB (Tsanas, 2012). All time series were *z*-standardized, and the phase space was normalized with the help of the Euclidean distance.

2.2.2.3. Transcription and coding. The lengthening within words, final pitch movements of intonation at the end of each phrase, pauses between and within words of a phrase, pauses between phrases, and accentuation were transcribed with the transcription editing software EXMARaLDA Partitur Editor (Version 1.4.4.; Schmidt & Wörner, 2009) while applying the GAT transcription system (Selting et al., 1998). The GAT transcription system is based on a variety of principles and conventions from various disciplines and provides guidelines for the notation of wording and prosody of spoken interaction (Selting et al., 1998). Based on the GAT transcription, we estimated the number of occurrences for lengthening, the inappropriate pauses within and between words of a phrase, the appropriate pauses between phrases, the accentuations (i.e., syllables that carried a focus accent), and the appropriate final pitch movements of intonation (i.e., falling intonations at the end of a sentence). All prosodic features were log transformed to reduce the skewness and then z-standardized.

3. Results

3.1. Data analysis

3.1.1. Recurrence quantification analysis

To capture the complexity and regularity of reading over time in each signal, we used *Recurrence Quantification Analysis* (RQA; Marwan et al., 2007; Wallot, 2017; Wallot et al., 2014) estimated with the CRP toolbox (Version 5.24 [R34]; Marwan, 2022) in the program MATLAB (Version R2017a; The Math Works, 2017). RQA is a general nonlinear time-series analysis technique (see also Wallot, 2017). The advantage of RQA compared to other methods is the possibility to quantify the complexity of a time series such as the extent that it is repetitive, noisy, or stationary. This quantification is accomplished through the

recurrence plot (Eckmann et al., 1987), which is the visualisation of the phase space trajectory of a dynamical system in a square matrix (see Fig. 3). The recurrence plot allows the extraction of a comprehensive set of measures. Four different recurrence plots for each child were generated: One for the amplitude of the signal, one for the duration of signal segments with amplitude above zero, one for the pitch of the signal, and one for the duration of all pauses in the signal. The parameters embedding dimension, delay and radius were held constant for each variable insuring that the average recurrence rate will be below 10 %. A complete list of the parameters can be found in the online supplement (see Table S1). The generated plots resulted in the extraction of 48 different RQA measures (12 variables per plot). Means and standard deviations for each signal were also estimated and included in the analyses. In the following text, we briefly describe the most important measures of RQA (Marwan & Webber, 2015; Webber & Zbilut, 2005) and provide an example for each measure in relation to amplitudes. A complete list of the measures can be found in the online supplement (see Table S2).

One of the most common measures of RQA is the *recurrence rate* (Marwan & Webber, 2015). This measure is the proportion of recurrence points to the total number of possible points in the recurrence plot (Wallot, 2017; see also Fig. 3). Therefore, it indicates the overall tendency of the signal to repeat itself within a specified radius (i.e., how often a state is revisited). This measure can range from 0 % to 100 %, with the lowest value indicating zero probability of recurrent points, whereas the highest value indicates that all points in the plot are recurrent. A higher recurrence rate in amplitudes, for example, would indicate an overall repetitiveness of amplitudes in the signal, i.e., readers insert a lot of stresses.

Another important measure, *percent determinism*, refers to the exact repetition of a sequence over time and is estimated by dividing the recurrence points of the diagonal lines by the total number of recurrence points in the recurrence plot (Wallot, 2017; see also Fig. 3). Similar to recurrence rate, the larger the value, the more exact repetitions in the signal. Put simply, percent determinism is a measure of the proportion of data points in the signal that are part of recurring patterns and indicates the predictability of the system (Marwan & Webber, 2015). Again, in relation to amplitudes, a higher determinism would indicate a more predictable system, with readers displaying more stress patterns over time. This might be the case if readers stress different words in the same way throughout the signal. Furthermore, recurrence rate and determinism of reading times have been shown to be positively related to reading skill (O'Brien et al., 2014), and determinism has been shown to be predictive of reading comprehension and reading speed (Wallot et al., 2014).

The third measure, *maximal diagonal line length*, refers to the length of the longest diagonal line in the recurrence plot apart from the main diagonal line (Wallot, 2017; see also Fig. 3). The inverse of this measure indicates the divergence of the trajectory segments. The diagonal lines in the recurrence plot of a chaotic dynamical system are short, whereas diagonal lines in the recurrence plot of a periodic system are longer. Therefore, the maximal diagonal line length of amplitudes reflects the duration of the signal's longest stress pattern.

The fourth measure, *entropy*, is a measure of the complexity of the

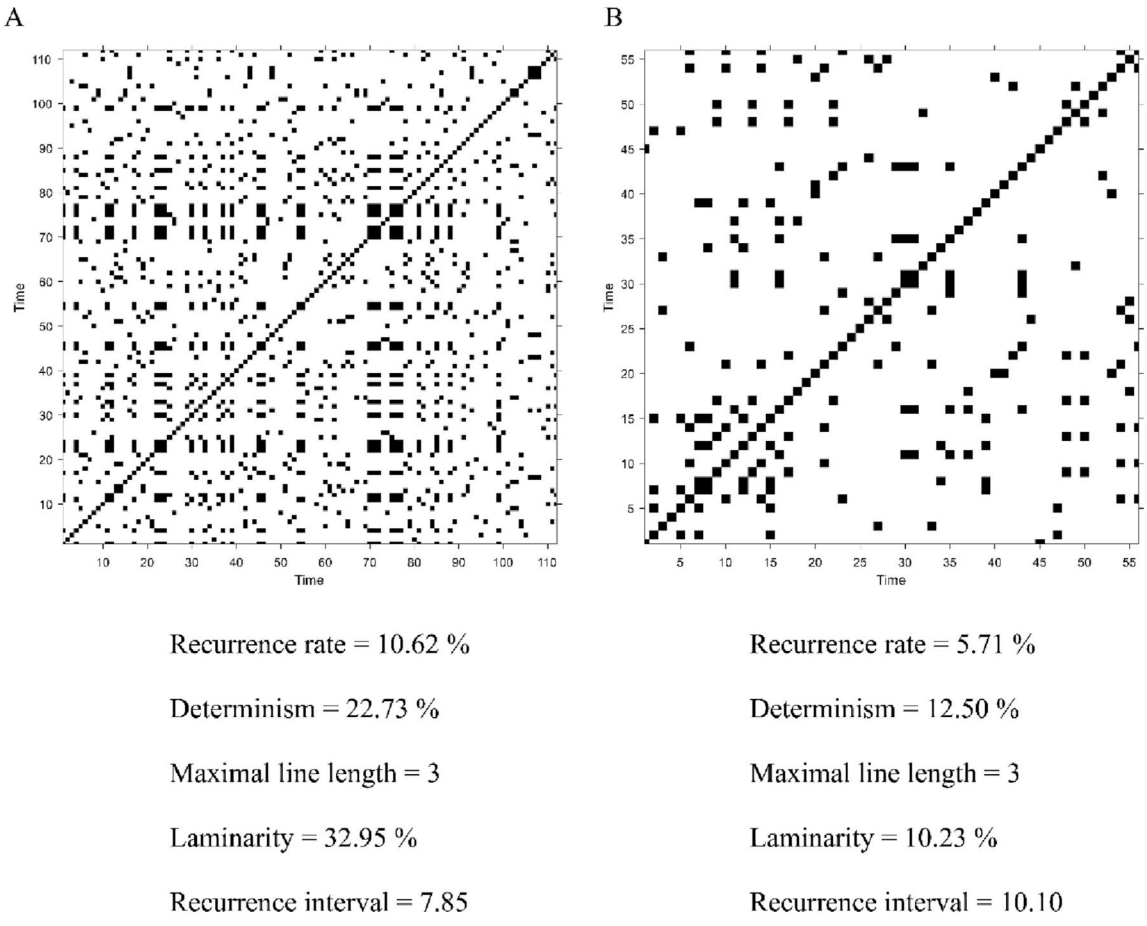


Fig. 3. Example of recurrence plots of amplitudes of (A) a struggling second grader and (B) a skilled fourth grader. Recurrence rate is estimated as the proportion of all recurrence points to all cells (without the main diagonal line); Determinism is estimated as the proportion of all diagonal recurrence points, parallel to the main diagonal, to all recurrence points (without the main diagonal line); Maximal line length is estimated as the longest non-main diagonal line; Laminarity is estimated as the proportion of all vertical recurrence points to all recurrence points; Recurrence interval is estimated by averaging the time needed until a recurrence is occurred.

signal and is computed as the Shannon's entropy of the frequency distribution of the lengths of diagonal lines (Marwan & Webber, 2015; see also Fig. 3). In short, entropy indicates the complexity of the patterns in the signal (Fusaroli & Tylén, 2016). The lower the entropy, the less complex and more predictable the system is (Leonardi, 2018), indicating the presence of fewer patterns (Fusaroli & Tylén, 2016). Consequently, a lower entropy of amplitudes indicates a more predictable system with fewer stress patterns, that is, stressing words in the same manner throughout the signal.

The fifth measure, *laminarity*, indicates the degree to which a signal is 'trapped' into repeating the same patterns over time (Cox et al., 2016; see also Fig. 3). Put it differently, laminarity quantifies the amount of smoothness (Konvalinka et al., 2011). Similar to determinism, laminarity is estimated by dividing the recurrence points of the vertical lines by the total number of recurrence points in the recurrence plot (Marwan et al., 2007). High laminarity indicates a smoother system that remains unchanged or changes slowly. The fewer vertical structures compared to single recurrence points, the lower the laminarity (Brick et al., 2018; Marwan et al., 2007). Accordingly, a high laminarity of amplitudes indicates that reader's stressing pattern changes at a slower rate. For example, this could be the case if readers stumble over words throughout the signal and stress the same letter several times in succession.

The sixth measure, *maximal vertical line length*, refers to the length of the longest horizontal and vertical line in the recurrence plot (Marwan et al., 2007). This line informs us about the maximal length of repetitions that are due to slow changing or identical consecutive occurrences (Leonardi, 2018). Therefore, the maximal vertical line length of amplitudes indicates the duration of the longest repetition of stresses in a row.

The seventh measure, *trend*, refers to the slope of the regression line of the amount of recurrence as a function of their distance from the main diagonal line (Marwan & Webber, 2015). Trend is an index of the degree of stationarity of the signal, with values near 0 indicating stationarity, whereas values far from 0 indicate a nonstationary process (Marwan & Webber, 2015). Hence, in relation to amplitudes a trend nearing zero indicates that the signal's stresses remain stationary over time, that is, they do not change over time.

Finally, *recurrence intervals* refer to the average time for the system to revisit a successive point in the signal (see Fig. 3). In other words, they reflect the average duration between recurrences and indicate the expected frequency of recurrences. Recurrence intervals are estimated by dividing the entire signal time by the total number of a recurrences. The shorter the recurrence interval, the more frequently a recurrence is likely to occur. Thus, a higher recurrence interval of amplitudes indicates a longer duration between recurring stresses.

3.1.2. Repeated k-fold cross-validation

Given the large number of variables obtained through RQA, we proceeded with a 100× repeated 5-fold cross-validation separately for each grade to train a LASSO (Least Absolute Shrinkage and Selection Operator) logistic regression model and identify the most important predictors for children's classification as skilled and struggling readers. To this end, we used the R caret package (Kuhn, 2008, 2020). In k-fold cross-validation, the data are partitioned in k subsets (folds), with k-1 folds used for the model construction and the one remaining fold used for the model validation (de Rooij & Weeda, 2020). In this study, the data were partitioned into five folds of approximately equal size. Four folds were randomly used to train the models and the one remaining fold was used to validate the final models. The exact same procedure was also used for the variables obtained through GAT.

Children's status as tutees or tutors (dummy-coded; 0 = tutee; 1 = tutor) was used as the outcome variable. All variables were z-standardized and entered as predictors in the models. The estimated parameters for each class were saved at the end of each calculation. These parameters were then used on the test data to obtain the prediction errors and the area under the curve. To increase the accuracy of the estimation and assess how many times a particular model had a smaller prediction error than the other models, this procedure was repeated 100 times. The built-in evaluation function of the caret package (Kuhn, 2008, 2020) was used to generate a list with the predictors ranked according to their contribution to the model (see Figs. S1–S4 in online supplement for the overall importance).

3.2. Important variables for distinguishing between tutees and tutors in grade 2

For the data of Grade 2, the 100× repeated 5-fold cross-validation showed that the model with the RQA variables had a larger area under the curve, a lower root mean square error, and a higher R^2 in the comparison of the two models (see Table 1). For the model with the RQA variables, the average reading rate per second ($B = 0.96$, $OR = 2.61$), the recurrence interval between pauses ($B = -0.49$, $OR = 0.61$), the maximal diagonal line length of amplitudes ($B = -0.17$, $OR = 0.84$), and the maximal vertical line length of pauses ($B = -0.01$, $OR = 0.99$) were the most important predictors for children's classification into tutees and tutors in Grade 2. For the model with the GAT variables, the inappropriate pauses within and between words of a phrase ($B = -0.75$, $OR = 0.47$) and the number of occurrences for lengthening ($B = -0.02$, $OR = 0.98$) were the most important predictors for children's classification into tutees and tutors in Grade 2.

3.3. Important variables for distinguishing between tutees and tutors in Grade 4

For the data of Grade 4, the 100× repeated 5-fold cross-validation showed that the model with the RQA variables had a lower area under the curve, but also a lower root mean square error, and a higher R^2 in the comparison of the two models (see Table 1). For the model with the RQA variables, a total of 20 predictors were found to be important for children's classification into tutees and tutors in Grade 4 (see Table 2 for the complete list of the predictors). The five most important predictors were the laminarity of pauses ($B = 3.65$, $OR = 38.44$), the average diagonal line of pauses ($B = 3.23$, $OR = 25.22$), the determinism of pitch ($B = -2.77$, $OR = 0.06$), the trend of recurrences in amplitudes ($B = -2.17$, OR = 0.11) and the maximal vertical line of pauses ($B = -2.01$, $OR = 0.13$). For the model with the GAT variables, the inappropriate pauses within and between words of a phrase ($B = -1.11$, $OR = 0.33$) and the number of occurrences for lengthening ($B = -0.12$, $OR = 0.89$) were the most important predictors for children's classification into tutees and tutors in Grade 4.

For each child in each model, logits were estimated and then transformed into probabilities. Estimated probabilities are presented as percentages. Classified struggling readers (tutees) from both models are shown in the bottom-left quadrant, whereas classified skilled readers (tutors) from both models are shown in the top-right quadrant.

4. Discussion

The primary goals of this study were to identify prosodic patterns in oral reading that differentiate struggling and skilled German readers using RQA measures, which have been successfully used to predict comprehension scores from reading time measures (Wallot et al., 2014), and to compare the discrimination of these patterns with prosodic features obtained using the transcription method GAT. For this purpose, oral readings of a passage by German second and fourth graders were recorded and analysed.

The results for Grade 2 revealed that the model estimated with the RQA measures had a better fit than the model estimated with the GAT measures. The average reading rate per second, the recurrence interval between pauses, the maximal diagonal line of amplitudes, and the maximal vertical line of pauses were the most important predictors of the model with the RQA variables.

Table 1
Comparison of Fit Indices in Models Estimated with 100× Repeated 5-fold Cross-Validation and Validated on the Test Fold with a Probability Cut-Off of 0.5.

Model	AUC	Sensitivity	Specificity	Accuracy	F1 score	RMSE	R^2
Grade 2							
LASSO model with RQA variables	0.96	0.90	0.77	0.84	0.85	0.40	0.47
LASSO model with GAT variables	0.83	0.70	0.67	0.68	0.70	0.56	0.13
Grade 4							
LASSO model with RQA variables	0.78	0.80	0.90	0.85	0.84	0.39	0.50
LASSO model with GAT variables	1.00	0.60	1.00	0.80	0.75	0.45	0.43

Note. AUC = area under the curve; Sensitivity = proportion of correct classified tutors; Specificity = proportion of correct classified tutees; Accuracy = accuracy of the model's prediction on the test fold; F1 score = mean between precision and recall; RMSE = root mean square error.

Table 2
Parameters and means of the most important predictors for classification of tutees and tutors by prosodic patterns in Grade 4.

Predictors	B	OR
Laminarity of pauses	3.65	38.44
Average diagonal line length of pauses	3.23	25.22
Determinism rate of pitch	−2.77	0.06
Linear trend of amplitudes	−2.17	0.11
Maximal vertical line of pauses	−2.01	0.13
Average diagonal line length of pronunciations	1.72	5.56
Average duration of pauses	−1.61	0.2
Maximal vertical line of amplitudes	−1.32	0.27
Recurrence interval between pitches	−1.25	0.29
Recurrence interval between amplitudes	−1.17	0.31
Entropy of amplitudes	−1.12	0.33
Entropy of pauses	0.85	2.34
Standard deviation of amplitudes	0.84	2.31
Laminarity of pitch	−0.72	0.49
Entropy of pitch	−0.69	0.5
Recurrence time entropy of pauses	0.44	1.55
Maximal diagonal line of pitch	−0.35	0.7
Trapping time of pitch	0.22	1.24
Recurrence interval between pauses	−0.14	0.87
Recurrence interval between pronunciations	0.02	1.02

The importance of the average reading rate in distinguishing between struggling and skilled readers should not be surprising considering that skilled readers have been shown to have a faster reading rate than struggling readers, reflecting more fluent reading (e.g., Schwanenflugel et al., 2004). Furthermore, one possible explanation for the longer intervals between recurring pauses could be that struggling readers stumble when reading a word or sentence, resulting in longer reading durations between pauses. This explanation, however, fails to account for the total number of pauses. According to the results of GAT, struggling readers still made 19 inappropriate pauses on average, compared to 5.71 inappropriate pauses on average for skilled readers. The third important variable, the maximal diagonal line, indicates that struggling readers have a more stable and periodic system regarding amplitudes than skilled readers. Struggling readers appear to have longer repetitions of amplitudes than skilled readers, which is consistent with prior research that revealed that skilled readers insert a stress less frequently than struggling readers (Clay & Imlach, 1971; Dowhower, 1991). According to Clay and Imlach (1971), good readers read in syntactic chunks, whereas poor readers read the text as if it was a list of words. Hence, it seems plausible that inserting a stress can result from inefficient word recognition. Readers who are unable to recognize words from memory, according to Torgesen and Hudson (2006), may stumble over several words and make numerous errors while trying to "sound out" the word before recognizing it and moving on to the next one. Finally, the maximal vertical line of pauses shows that struggling readers have a longer sequence of consecutive pauses than skilled readers. The predictor "pauses" includes intra- and inter-sentential pauses, that is, pauses within and between words independent of the syntactic phrases. This finding is in line with prior research, which found that skilled readers take fewer pauses than struggling readers (e.g., Clay & Imlach, 1971; Schwanenflugel et al., 2004).

In Grade 4, the results revealed that again the model estimated with the RQA measures had a better fit than the model estimated with the GAT measures. A total of 20 variables were found to be important in classifying tutees and tutors. However, in this section, we will focus on the five most important predictors: laminarity of pauses, average diagonal line of pauses, the determinism of pitch, trend of recurrences in amplitudes, and maximal vertical line of pauses.

It appears that pause patterns of skilled readers change at a slower pace than those of struggling readers, indicating a higher degree of repeating pause patterns. Additionally, our results suggest that skilled readers show on average more repetitions of a pause pattern than struggling readers. However, these findings do not imply that skilled readers took more or longer pauses than struggling readers. It simply shows that the pause patterns of skilled readers were more stable over time than the pause patterns of struggling readers. Furthermore, determinism of pitch suggests that struggling readers exhibit more pitch repetitions than skilled readers. One possible explanation for the higher pitch repetitiveness could be that struggling readers read words or a whole phrase with the same pitch, compared to skilled readers, who have been shown to have a larger pitch range and more appropriate pitch falls and rises than struggling readers (e.g., English: Benjamin et al., 2013; Spanish: Álvarez-Cañizo et al., 2015). In addition, when compared to skilled readers, struggling readers showed similar amplitude patterns over time. This finding is consistent with previous studies which showed that struggling readers insert a stress in each word, whereas skilled readers insert a stress less frequently (Clay & Imlach, 1971). Hence, amplitudes of struggling readers are distributed more homogenously over time than the amplitudes of skilled readers. Finally, the maximal vertical line of pauses shows that struggling readers have a longer sequence of consecutive pauses than skilled readers. This result is also in accordance with previous studies, which showed that skilled readers make fewer pauses compared to struggling readers (e.g., Clay & Imlach, 1971; Schwanenflugel et al., 2004).

According to the results of the prosodic features obtained with coding schema GAT, the sum of inappropriate pauses and occurrences of lengthening were the most important predictors in distinguishing between skilled and struggling readers in Grades 2 and 4. Struggling readers made hesitant pauses within words or in the middle of a phrase. Skilled readers making fewer and shorter ungrammatical pauses than struggling readers is also consistent with prior studies (Álvarez-Cañizo et al., 2015; Godde et al., 2020; Schwanenflugel et al., 2004). Inappropriate pauses appear to be the result of decoding issues or the misreading of a sentence's syntax (Godde et al., 2020; Miller & Schwanenflugel, 2008) and have been shown to negatively relate to other aspects of reading fluency and reading comprehension (Benjamin & Schwanenflugel, 2010; Dowhower, 1991). Furthermore, struggling readers in the current study prolonged syllables within words, which lead to inappropriate word pronunciation. This finding is in line with Lyytinen et al. (1995), who showed that lengthening occurrences within words in struggling Finnish readers occurs more frequently than in skilled readers. One possible explanation for this improper use of vowel duration marking could be that children have difficulty distinguishing between long and short vowel phonemes (Landerl & Reitsma, 2005).

In accordance with our expectations, when models with RQA predictors were compared to models with GAT predictors, models with RQA predictors had a higher area under the curve (except for the area under the curve in Grade 4), a smaller root mean square error, and a higher R^2 of the two models than the GAT models. The models estimated using prosodic patterns performed better than models estimated with prosodic features. However, as can be seen in the upper left and lower right quadrants of Figs. 4 and 5, the children misclassified by the two methods were usually not the same. This misclassification might be caused by children demonstrating a behavior pattern in the variables of interest that differs from their respective group. For example, by looking at the incorrect pauses of second grade tutors who were misclassified as tutees using the GAT model and the incorrect pauses of correctly classified second grade tutors, we noticed that misclassified tutors made 11.80 incorrect pauses on average, whereas correctly classified tutors made only 4.15 incorrect pauses on average. Nevertheless, it appears that despite tapping into distinct processes, both methods can distinguish between skilled and struggling readers, with complex prosodic patterns obtained with RQA faring better than prosodic features obtained with GAT.

In sum, our findings indicate that German skilled readers in Grade 2 have a faster reading rate, shorter intervals between pauses, and less repetitions of recurrent amplitudes and pauses, whereas German skilled readers in Grade 4 had more stable pause patterns over time, less pitch repetitions, less similar amplitude patterns over time, and less repetitions of pauses in a row. Furthermore, the RQA models outperformed the GAT models in the classification of skilled and struggling readers.

4.1. Limitations

Given the scarcity of RQA for prosody data, our findings are encouraging. They must, however, be interpreted with some limitations in mind. First, because of the large number of variables, we needed to conduct a 100 × 5-fold cross-validation to deal with the RQA's curse of dimensionality. We reduced this number by employing a LASSO logistic regression to identify the most important predictors influencing model performance with a feature selection algorithm. Nonetheless, despite the exploratory nature of our analyses, this procedure may alleviate some concerns regarding overfitting. Therefore, we made every effort to document and disclose every stage of our analyses so that they may be utilized as a reference for future studies.

Another limitation of our study is that even though children were separated into tutees and tutors at the start of the study based on their pre-test reading comprehension scores, they had already participated in 23 training sessions by the time of the oral reading. Therefore, notwithstanding the disparities in reading comprehension between the two groups at the post-test, we cannot be certain that the results would have been the same if the children had not participated in the training.

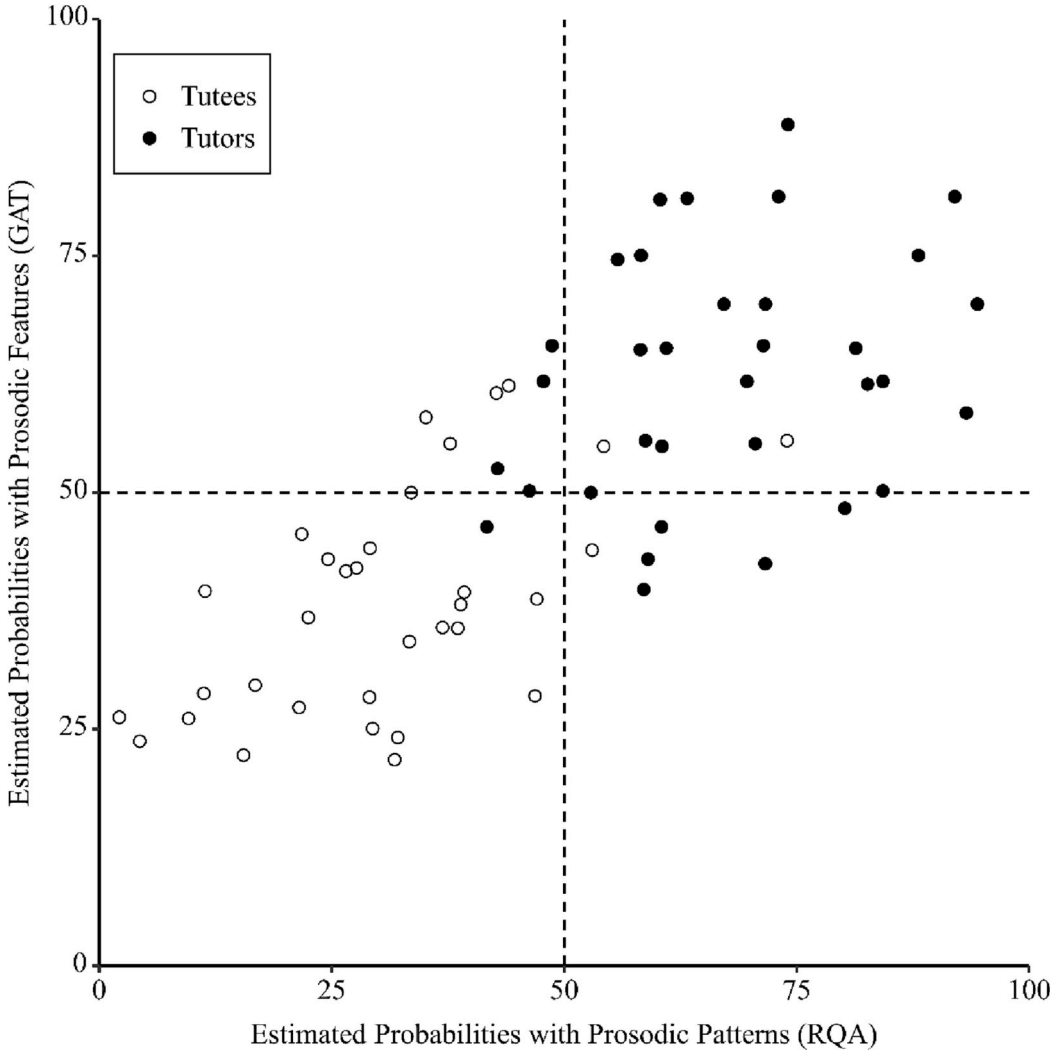


Fig. 4. Scatter plot of the estimated probabilities from the model with prosodic features versus the estimated probabilities for the model with prosodic patterns in Grade 2.

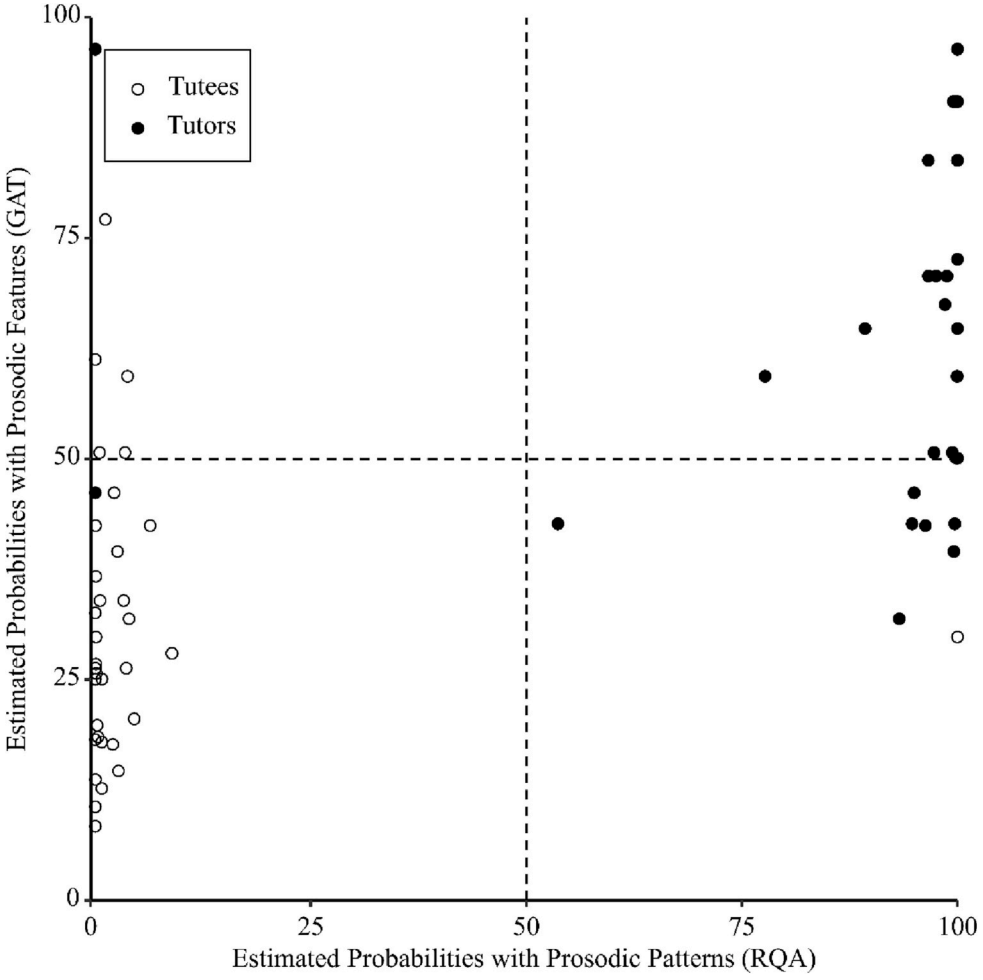


Fig. 5. Scatter plot of the estimated probabilities from the model with prosodic features versus the estimated probabilities for the model with prosodic patterns in Grade 4. For each child in each model, logits were estimated and then transformed into probabilities. Estimated probabilities are presented as percentages. Classified struggling readers (tutees) from both models are shown in the bottom-left quadrant, whereas classified skilled readers (tutors) from both models are shown in the top-right quadrant.

In addition, half of the children received a reading fluency training whereas the other half received a training of visuospatial working memory. However, investigating prosodic patterns separately for each training would require a larger sample. Sample sizes in this study were relatively small, which might have also impacted the parameter estimates and lead to a relatively low power.

4.2. Conclusion

Despite these limitations, the study demonstrates that using RQA for prosody analysis is a promising approach. If similar results were to be obtained in a longitudinal design identifying prosodic patterns that distinguish between struggling and skilled readers throughout primary school, RQA might prove to be a versatile tool in research on prosody. Confirming our findings would provide experts with another option for a precise semi-automatic assessment of prosodic difficulties, which is far more economic than hand-coding prosodic features with GAT. Finally, the RQA seems to provide additional information about prosody that augments established approaches, which might be useful for advancing the existing theories.

Funding

The research reported in this article was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under Grant 01GJ1004. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. The protocol was approved by the "Ministry of Education and Cultural Affairs, Hesse" (cf. Education Act of Hesse, Section 84).

Consent to participate

This study was carried out in accordance with the recommendations of "Ministry of Education and Cultural Affairs, Hesse, Germany (Hessisches Kultusministerium)" with written informed consent from all subjects. The parents of all subjects gave written informed consent in accordance with the Declaration of Helsinki. Principals of the schools participating in the study gave written consent after the school conference (i.e., the majority of teachers agreed to realize the study in their school).

Declaration of competing interest

This article is based on data published in Müller et al. (2015). We have no known conflict of interest to disclose.

Data availability

All data that support the findings of this study are available in the repository of the Open Science Framework (OSF) at https://osf.io/ep4bw/?view_only=f44d044d7bce4030b50a7c6e41f17ce6.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2023.103892.

References

Allington, R. L. (1983). Fluency: The neglected reading goal. *The Reading Teacher, 36*(6), 556–561.

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112*(1), 22–39. https://doi.org/10.1037/edu0000359

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2015). The role of reading fluency in children's text comprehension. *Frontiers in Psychology, 6*, Article 1810. https://doi.org/10.3389/fpsyg.2015.01810

Audacity Team. (2006). *Audacity: Free audio editor and recorder (version 1.2.6) [computer software]*. Audacity Team. https://audacityteam.org.

Bailly, G., & Gouvernayre, C. (2012). Pauses and respiratory markers of the structure of book reading. In *13th annual conference of the international speech communication association 2012 (INTERSPEECH 2012)* (pp. 2218–2221). Curran Associates. https://doi.org/10.21437/Interspeech.2012-591.

Bamberger, R., & Rabin, A. T. (1984). New approaches to readability: Austrian research. *The Reading Teacher, 37*(6), 512–519.

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45*(4), 388–404. https://doi.org/10.1598/RRQ.45.4.2

Benjamin, R. G., Schwanenflugel, P. J., Meisinger, E. B., Groff, C., Kuhn, M. R., & Steiner, R. (2013). A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly, 48*(2), 105–133. https://doi.org/10.1002/rrq.43

Björnsson, C. H. (1968). *Läsbarhet [Readability]*. Liber.

Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer program]. *Phonetic Sciences*. https://www.fon.hum.uva.nl/praat/.

Bressem, J. (2013). Transcription systems for gestures, speech, prosody, postures, and gaze. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body – Language – Communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 1037–1059). De Gruyter Mouton. https://doi.org/10.1515/9783110261318.1037.

Brick, T. R., Gray, A. L., & Staples, A. D. (2018). Recurrence quantification for the analysis of coupled processes in aging. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 73*(1), 134–147. https://doi.org/10.1093/geronb/gbx018

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*(1), 108–132. https://doi.org/10.1006/jmps.1999.1279

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel, & I. Marlien (Eds.), *Speech prosody 2002: Proceedings: Aix-en-Provence, France, 11–13 April 2002* (pp. 199–202). Laboratoire Parole et Langage.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Clay, M. M., & Imlach, R. H. (1971). Juncture, pitch, and stress as reading behavior variables. *Journal of Verbal Learning & Verbal Behavior, 10*(2), 133–139. https://doi.org/10.1016/S0022-5371(71)80004-X

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cooper, W. E., & Cooper, J. (1980). *Syntax and speech*. Harvard University Press.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Edward Arnold.

Cox, R. F. A., van der Steen, S., Guevara Guerrero, M., Hoekstra, L., & van Dijk, M. (2016). Chromatic and anisotropic cross-recurrence quantification analysis of interpersonal behavior. In C. Webber, C. Ioana, & N. Marwan (Eds.), *Recurrence plots and their quantifications: Expanding horizons: Proceedings of the 6th international symposium on recurrence plots, Grenoble, France, 17-19 June 2015* (pp. 209–225). Springer. https://doi.org/10.1007/978-3-319-29922-8.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006–469). *National Center for Education Statistics*. https://files.eric.ed.gov/fulltext/ED488962.pdf.

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science, 3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22*(4), 389–406. https://doi.org/10.2307/747699

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice, 30*(3), 165–175. https://doi.org/10.1080/00405849109543497

Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters, 4*(9), 973–977. https://doi.org/10.1209/0295-5075/4/9/004

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America, 89*(1), 369–382. https://doi.org/10.1121/1.400674

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science, 40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Godde, E., Baily, G., & Bosse, M.-L. (2022). Pausing and breathing while reading aloud: Development from 2nd to 7th grade in french speaking children. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10168-z

Godde, E., Bosse, M.-L., & Bailly, G. (2020). A review of reading prosody acquisition and development. *Reading and Writing, 33*, 399–426. https://doi.org/10.1007/s11145-019-09968-1

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics, 71*(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Himmelmann, N. P., & Ladd, D. R. (2008). Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation, 2*(2), 244–274.

Hirst, D. (1991). Intonation models: Towards a third generation. In , *1. Proceedings of the 12th International Congress of Phonétic Sciences* (pp. 305–310).

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly, 25*(1), 4–32. https://doi.org/10.1080/10573560802491208

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology, 106*(4), 1096–1106. https://doi.org/10.1037/a0037100

Kara, Y., Kamata, A., Potgieter, C., & Nese, J. F. T. (2020). Estimating model-based oral reading fluency: A Bayesian approach. *Educational and Psychological Measurement, 80*(5), 847–869. https://doi.org/10.1177/0013164419900208

Karageorgos, P., Müller, B., & Richter, T. (2019). Modelling the relationship of accurate and fluent word recognition in primary school. *Learning and Individual Differences, 76*, 1–11. https://doi.org/10.1016/j.lindif.2019.101779

Karageorgos, P., Richter, T., Haffmans, M.-B., Schindler, J., & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development, 56*, 1–13. https://doi.org/10.1016/j.cogdev.2020.100949

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., Van Orden, G., & Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108*(20), 8514–8519. https://doi.org/10.1073/pnas.1016955108

Ktori, M., Mousikou, P., & Rastle, K. (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology: General, 147*(1), 36–61. https://doi.org/10.1037/xge0000380

Kuhn, M. (2008). Building predicitve models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. (2020). *caret: Classification and regression training. R package (Version 6.0-86) [Computer software]*. https://CRAN.R-project.org/package=caret.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in dutch and german. *Journal of Experimental Child Psychology, 92*(4), 322–344. https://doi.org/10.1016/j.jecp.2005.04.005

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Lenhard, W., & Schneider, W. (2006). *ELFE 1–6: Ein Leseverständnistest für Erst- bis Sechstklässler [ELFE 1-6: A reading comprehension test for children in grades one to six]*. Hogrefe.

Leonardi, G. (2018). A method for the computation of entropy in the recurrence quantification analysis of categorical time series. *Physica A: Statistical Mechanics and its Applications, 512*, 824–836. https://doi.org/10.1016/j.physa.2018.08.058

Lyytinen, H., Leinonen, S., Nikula, M., Aro, M., & Leiwo, M. (1995). In search of the core features of dyslexia: Observations concerning dyslexia in the highly orthographically regular finnish language. In V. Berninger (Ed.), *The varieties of orthographic knowledge II: Relationships to phonology, reading, and writing* (pp. 177–204). Kluwer.

Marwan, N. (2022). *Cross recurrence plot toolbox for MATLAB (Version 5.24[R34]) [Computer software]*. https://tocsy.pik-potsdam.de/CRPtoolbox/.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Marwan, N., & Webber, C. L., Jr. (2015). Mathematical and computational foundations of recurrence quantifications. In N. Marwan, & C. L. Webber, Jr. (Eds.), *Recurrence quantification analysis: Theory and best practices* (pp. 3–42). Springer. https://doi.org/10.1007/978-3-319-07155-8.

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in german learners of english. *Studies in Second Language Acquisition, 36*(2), 303–329. https://doi.org/10.1017/S0272263114000023

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of english and german. *The Journal of*

the Acoustical Society of America, 131(3), 2249–2260. https://doi.org/10.1121/1.3681950
Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology, 98*(4), 839–843. https://doi.org/10.1037/0022-0663.98.4.839
Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly, 43*(4), 336–354. https://doi.org/10.1598/RRQ.43.4.2
Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities, 24*(1), 81–92. https://doi.org/10.1007/BF00115030
Morrison, T. G., & Wilcox, B. (2020). Assessing expressive oral reading fluency. *Education Sciences, 10*(3), 59. https://doi.org/10.3390/educsci10030059
Müller, B., Mayer, A., Richter, T., Križan, A., Hecht, T., & Ennemoser, M. (2015). Differential effects of reading trainings on reading processes: A comparison in Grade 2. *Zeitschrift für Erziehungswissenschaft, 18*, 489–512. https://doi.org/10.1007/s11618-015-0648-0
National Institute of Child Health and Human Development. (2000). Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. In *NIH Publication No. 00-4769*. U.S. Government Printing Office. https://files.eric.ed.gov/fulltext/ED444126.pdf.
O'Brien, B. A., Wallot, S., Haussmann, A., & Kloos, H. (2014). Using complexity metrics to assess silent reading fluency: A cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading, 18*(4), 235–254. https://doi.org/10.1080/10888438.2013.862248
Paige, D. D., Rasinski, T., Magpuri-Lavell, T., & Smith, G. S. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research, 46*(2), 123–156. https://doi.org/10.1177/1086296X14535170
Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). Listening to children read aloud: Data from NAEP's integrated reading performance record (IRPR) at grade 4. In *NCES publication no. 95-726*. National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED378550.pdf.
Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading and Writing, 19*(2), 171–198. https://doi.org/10.1007/s11145-005-4023-z
Rasinski, T. (2004). Creating fluent readers. *Educational Leadership, 61*(6), 46–51. http://educationalleader.com/subtopicintro/read/ASCD/ASCD_364_1.pdf.
Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350–361. https://doi.org/10.1080/19388070802468715
Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language, 42*(3), 342–364. https://doi.org/10.1006/jmla.1999.2687
Scharff-Rethfeldt, W., Miller, N., & Mennen, I. (2008). Unterschiede in der mittleren Sprechtonhöhe bei Deutsch/Englisch bilingualen Sprechern [Speaking fundamental frequency differences in German-English bilinguals]. *Sprache Stimme Gehör, 32*(3), 123–128. https://doi.org/10.1055/s-0028-1083799
Schmidt, T., & Wörner, K. (2009). EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics, 19*(4), 565–582. https://doi.org/10.1075/prag.19.4.06sch
Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*(1), 119–129. https://doi.org/10.1037/0022-0663.96.1.119
Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kulhen, E., Günther, S., Quasthoff, U., Meier, C., Schlobinski, P., & Uhmann, S. (1998). Gesprächsanalytisches transkriptionssystem (GAT) [Discourse and conversation-analytic transcription system]. *Linguistische Berichte, 173*, 91–122.
Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P. J., Pierrehumbert, J. B., & Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In *Proc. 2nd international conference on spoken language processing (ICSLP 1992)* (pp. 867–870).
Silverman, R. D., Speece, D. L., & Harring, J. R. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading, 17*(2), 108–133. https://doi.org/10.1080/10888438.2011.618153
Stevens, K. C. (1981). Chunking material as an aid to reading comprehension. *Journal of Reading, 25*(2), 126–129.
Temperley, D. (2009). Distributional stress regularity: A corpus study. *Journal of Psycholinguistic Research, 38*(1), 75–92. https://doi.org/10.1007/s10936-008-9084-0
The Math Works, I. (2017). *MATLAB (version R2017a) [computer software]*. MathWorks. https://www.mathworks.com/.
Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America, 85*(4), 1699–1707. https://doi.org/10.1121/1.397959
Torgesen, J. K., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling readers. In S. J. Samuels, & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 130–158). International Reading Association.
Tsanas, A. (2012). *Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning*. Oxford University. Unpublished doctoral dissertation.
Tschense, M., & Wallot, S. (2022). Using measures of reading time regularity (RTR) to quantify eye movement dynamics, and how they are shaped by linguistic information. *Journal of Vision, 22*(9), 1–21. https://doi.org/10.1167/jov.22.6.9
Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001
Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes, 54*(5–6), 382–405. https://doi.org/10.1080/0163853X.2017.1297921
Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 40*(6), 1745–1765. https://doi.org/10.1037/xlm0000030
Webber, C. L., & Zbilut, J. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley, & G. C. V. Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–94). https://www.nsf.gov/pubs/2005/nsf05057/nmbs/nmbs.pdf.
Wolters, A. P., Kim, Y.-S. G., & Szura, J. W. (2020). Is reading prosody related to reading comprehension?A meta-analysis. *Scientific Studies of Reading, 1–20*. https://doi.org/10.1080/10888438.2020.1850733