Modelling the eye movements of dyslexic children during reading as a Continuous Time Random Walk

F. R. Iaconis,[1, a)] M. Meo,[1] J. A. Del Punta,[2] and G. Gasaneo[3]

[1)] *Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur (UNS) - CONICET, 8000 Bahía Blanca, Argentina*

[2)] *Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur (UNS) - CONICET and Departamento de Matemática, Universidad Nacional del Sur (UNS), 8000 Bahía Blanca, Argentina*

[3)] *Centro Integral de Neuricencias Aplicadas, 8000, Bahía Blanca, Argentina, and Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur (UNS) - CONICET, 8000 Bahía Blanca, Argentina*

(Dated: 10 May 2023)

The study of eye movements during reading is considered a valuable tool for understanding the underlying cognitive processes and for its ability to detect alterations that could be associated with neurocognitive deficiencies or visual conditions. During reading the gaze moves from one position to the next on the text performing a saccade–fixation sequence. This dynamics resembles processes usually described as Continuous Time Random Walk, where the jumps are the saccadic movements, and the waiting times are the duration of fixations. The time between jumps (intersaccadic time) consists of a stochastic waiting time and a flight time which is a function of the jump length (the amplitude of the saccade). This motivates the present proposal of a model of eye movements during reading in the framework of the Intermittent Random Walk, but considering the time between jumps as combined stochastic-deterministic process. The parameters used in this model were obtained from records of eye movements of children with dyslexia and typically developed children performing a reading task. The jump lengths arise from the characteristics of the selected text. The time required for the flights was obtained based on a previously proposed model. Synthetic signals were generated and compared with actual eye movements signals in a complexity-entropy plane.

Dyslexia is a neurodevelopmental disorder responsible for one of the most frequent learning disabilities characterised by difficulties in reading. The eye movements of individuals with dyslexia present differences with respect to those of typically developed individuals during reading.[1–3]. For this reason, the study of eye movements of children with dyslexia has become a topic of great interest in different scientific communities, motivated by the idea that a deep understanding of them allows early diagnosis and prompt intervention in the process of learning to read. As a contribution to this topic, we present a simple model of eye movements of children with dyslexia based on the Continuous Time Random Walk framework, taking into account the combination of stochastic and deterministic processes that characterise eye movements during reading. The model involves few parameters that can be derived from a small number of experimental data. Such simplicity has the advantage of allowing the generation of synthetic eye movement data in a simple way that could be of interest for studies using techniques that require large amounts of data that are not easy to obtain.

I. INTRODUCTION

The study of eye movements during reading has a very long history. In the last 20 years, researches in this area have focused on the development of models that account for the different mechanisms involved in the task[4]. Two fundamental principles govern the dynamics of reading: *where* to look and *when* to move the eyes to the next target. In an attempt to explain them, the developed models include a large number of elements that allow the reading process to be described in great detail[4,5]. The most common elements usually included are fixation duration, saccade length, processing time (where first fixation duration, single fixation duration or gaze duration are intertwined), skipped words, and regressions[4]. The more elements are included in the model, the more interactions must be taken into account. Thus, not only does the complexity of the algorithms increases, but the descriptions become dependent on many specific details, which results in the loss of the ability to simply describe the phenomenon in general, beyond the particular details involved in each sub part of the process.

Globally, the reading process consists of a sequence of steps that varies depending on the skills of the reader and the type of reading that the subject performs[6]. Once a person acquires some reading skills, differences in strategies become more relevant. These strategies can be related to the particularities of the text or the reading task, but also to alterations that affect the visual system or certain brain areas involved in reading.

Particularly, we are interested in dyslexia, a condition detected in children mostly during schooling, that originate a particular way of processing and reading. Several researches address the study of eye movements of readers with dyslexia[1–3,7–9], following the premise that an adequate description of them can lead to an early diagnosis and intervention. Among the various existing models of eye movements during reading, some could be adapted to include the mod-

[a)]Electronic mail: franciscoiaconis@gmail.com

elling of readers with dyslexia. However, as previously mentioned, these models involve a large number of elements and interactions that make them very specific and complex.

A simple model, capable of globally describing reading with a reduced number of parameters, and allowing to recognise the specific elements that characterise both strategies and particularities of the readers, may be useful in many aspects. One of them could be its usability to provide simulated data sets which can be of great interest in studies that require large volumes of data, or data that is not simple to obtain. Alternatively, it may be useful to experiment with tools that can lead to new forms of diagnosis by allowing the detection of quantifiable features. For example, an elementary model of saccadic eye movements[10] has been implemented to generate synthetic signals that were used to study event detection and denoising techniques[11]. In the same way, in the present work we propose a model that allows generating synthetic eye movement signals of dyslexic children during reading in a simple and adequate way that could be used in further studies or even being useful as a clinical tool for diagnostic purposes.

Our proposal combines three main theoretical approaches. The first one refers to the approach from which we address the reading process. From a physically observable point of view, reading is directly related to the stochastic dynamics of the eyeball (a rigid body) where a sequence of rotations and stops are observed. Such a behaviour can be related to a typical process described by the Continuous Time Random Walk (CTRW) theory[12–15]. This framework is an extension of the classical Brownian motion where the waiting times and jump lengths are obtained from probability distributions that, coupled or separable, can lead to sub-diffusive or super-diffusive processes. Different diffusive processes have been modelled with the CTRW [16,17]. In the framework of the CTRW and based on the Master equation, Gomez Portillo and collaborators[14] have discussed a process where the jumps, also called flights, were characterised by random speeds with a certain duration (flight periods). In the present manuscript we discuss an extension of the proposal of Gomez Portillo, but instead of using master equations we implement a set of subordinate equations as done by, e.g., Fogedby and collaborators[18]. In our model, the flights (saccadic movements) are modelled following some equations presented in a previous works[19,20] where the flight duration is a function of the saccade amplitude.

The second key element in our proposal refers to the choice of the probability density functions used to model the duration of fixations and the amplitude of saccades. We propose linear combinations of lognormal functions to this end, a choice that results from analysing previous publications, where eye movements performed during cognitive tasks were characterised as nonlinear processes[15,21–25]. According to these approaches, reading is a cognitive process in which different components interact to produce a response to a certain stimulus. Many cognitive researchers have studied cognitive processes focusing on the characterisation of components and interactions, distinguishing apart processes component-dominated or interaction-dominated[21,23,26,27]. In particular, Stephen and coworkers[21,23] have shown that the distribution of gaze steps (the distance from one sampled point to the next in an eye tracker dataset) can be modelled with lognormal or power law distributions that describes multiplicative processes with interaction-dominant structures. Also lognormal functions have been proposed to model fixation durations and saccade amplitudes[28] in a relatively simple proposal based on mixture models.

The third element refers to the complexity of the components and interactions involved in cognitive tasks. Complexity, understood as an intermediate state between order and disorder and related to scale-free structures, has been used to describe different biological systems, particularly neural dynamics[29]. Hidden structures of selected features can be studied using statistical complexity measures[30,31]. Defined as the product of the Shannon entropy and a discrete version of the Jensen-Shannon divergence, the complexity index organises the structures from a minimum value, indicating a fully ordered or disordered system, to a maximum, reached in an intermediate value. When plotted against Shannon entropy one obtains a complexity-entropy plane that provides information about the structures of the studied patterns in terms of their randomness. Recently, the complexity-entropy plane has been used to characterise cognitive differences in a psychological test that assesses attention, from the analysis of different patterns emerging from eye movements dynamics[25]. This tool is robust enough to separate different cognitive or behavioural characteristics. Thus, complexity-entropy plots can be used not only to characterise the hidden structures of selected features of eye movements during reading, but also to test the proposed model by comparing the Shannon entropy and statistical complexity of actual eye movement signals registered from dyslexic children during reading and synthetic data generated with an algorithm based on the dynamic approach we propose here.

This presentation is organised as follows. We start with a brief description of eye movement dynamics of dyslexic readers and a summary of the experiment conducted and data collection. In Sec. III we characterise the four fundamental features involved in the model we propose: duration of fixations, and amplitude, direction and duration of saccades. The formalism of the model and an algorithm to generate synthetic data are presented in Sec. IV. A set of simulated signals are analysed and compared to actual eye movement records in Sec. V. The complexity-entropy profiles of fixations and saccades are also studied. Finally, a discussion and some conclusions close our proposal in Sec. VI.

II. EYE MOVEMENT DYNAMICS DURING READING

A. Reading and dyslexia

Dyslexia is a specific learning disability classified as a neurodevelopmental disorder by the American Psychiatric Association[32]. Affecting around 10% of the population[33], dyslexia is considered the most frequent learning disability and is characterised by various difficulties in reading like, e.g., decoding and writing words even with an otherwise adequate intellectual development[33–35].

During reading, eye movements of dyslexic individuals differ from those of typically developed (TD) readers, independently of the languages considered[1–3]. The observed differences seem to be related to difficulties of dyslexic individuals to consolidate the necessary automation of the grapheme-phoneme transformation. This process develops smoothly in TD children, but does not seem to occur easily in dyslexic children[2]. Previous studies performed on English and other non-transparent language speakers, established that during reading, dyslexic readers present longer duration of fixations, shorter saccades and more regressions than TD readers of the same chronological age[3,7,8]. This correlates with the strategy of collecting the information presented in the text piece by piece, relating each grapheme to the known phoneme.

B. Participants and experimental data

12 dyslexic children aged 9-10 years compose the group of participants on which we focus our proposal. The dyslexia diagnosis was performed by psychopedagogues specialised on reading disorders. For comparison purposes, a group of 29 children in the same age range was also included in the study.

All participants speak Spanish as native language, and attend medium socioeconomical level schools in the Buenos Aires Metropolitan Area (Argentina). Children were asked to read aloud a short Spanish nine-line text shown on a computer monitor while their eye movements were recorded, at a sample rate of 90 Hz, using the eye tracker Tobii Pro (Tobii AB, Sweden). The records of dyslexic children were registered in psycho-pedagogical clinics. The eye movements of the TD readers were recorded by professionals in the school they attend. Informed consent was obtained from the legal guardian of each participant. The children also gave their consent to participate and the school authorised the collection of data within the institution. All participants were treated in full accordance with the Declaration of Helsinki.

Each collected dataset constitutes a two dimensional time series $\{(t,x,y)\}$ indicating the eye position (x,y) on the screen at each sampled time t in milliseconds (ms). Positions on the screen were normalised to satisfy the ratio 16:9. Thus, in what follows, x ranges from 0 to 16 and y ranges from 0 to 9. Fixations and saccades were detected with an algorithm adapted from the one proposed in Ref.[36]. Blinks and saccades associated with return-sweeps (saccades that take the gaze from the end of one line to the beginning of the next) were eliminated from the analysis.

For a correct statistical analysis large amounts of eye movements are needed. In the present study, despite the fact that the number of participants is relatively low, a reliable statistical analysis can be performed since each participant performs around 200 movements to read the text used in our experiment. The analysis we carried out was made with a total of 2130 saccadic movements and fixations.

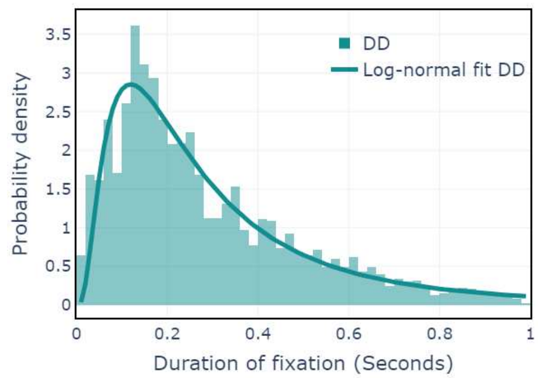


FIG. 1. Histogram of duration of fixations of dyslexic children, obtained from the experimental data. In full line the lognormal fit.

III. ANALYSIS OF EYE MOVEMENT FEATURES DURING READING

To model eye movements, information about the duration of fixations and the duration and amplitude of saccades is required. In this section we analyse these features based on the experimental data collected in order to establish the basic elements to be included in the model.

A. Fixations

The duration of fixations is generally modulated by the underlying cognitive process and varies considerably throughout it. Different type of probability density functions have been used to represent the occurrence of fixation durations, among the most extensively used is the lognormal function[28]. Fig. 1 shows the histogram of the duration of fixations obtained form the dyslexic children. Frequencies can be fitted with a lognormal function:

$$f(\mu,\sigma;x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}} \tag{1}$$

which can be used as the probability density function (PDF) of the duration of fixations. We performed this fit using the curve_fit method included in the scipy.optimize python package. The values obtained for the parameters were: $\mu = -1.44 \pm 0.04$ and $\sigma = 0.83 \pm 0.03$. The resulting curve is plotted in Fig. 1.

In order to verify the adequacy of these values, the procedure was repeated 6 times but taking half of the records randomly selected finding in all cases similar values for μ and σ. The procedure was also implemented with the data collected from the TD children. The histogram of duration of fixations for this group did not present significant differences with respect to the dyslexic ones.

B. Amplitude and direction of saccades based on a stimulus

Learning to read in Spanish requires, as a first step, recognising the letters. Next, syllables or graphemes and their corresponding phonemes are identified. The association between graphemes and phonemes must become automatic to continue with the following step: the recognition of the word as a whole, formed from a given sequence of graphemes. While consolidating reading, children gradually incorporate words to this automatising process. The known words are transformed immediately, the unknown words are decomposed into graphemes and transformed to phonemes. Thus, depending on how consolidated the process is, it will be how reading will be carried out, that is, how much mix there will be between word recognition and syllable processing. Dyslexic children present difficulties in the automation process, which in many cases never occur[2,37].

The movement of the eyes over the text clearly depends on how automated the reading process is. This suggests that the analysis of the distances between syllables and words in the text is critical in understanding the dynamics of eye movements during reading, especially in relation to the level of reading proficiency. The spatial configuration of stimuli will strongly influence the amplitude of saccades. Less skilled readers, whether dyslexic or slow learners, will perform saccadic movements with amplitudes close to the distances between consecutive syllables in the text. As the reader becomes more skilled, it is expected that short words of up to two syllables will be recognised as a whole and the amplitude of saccades will be closer to the distance of two syllables[38,39]. We call this distance second neighbours distance, and we call first neighbours distance the distance between consecutive syllables. More experienced readers do not need to fixate on words of less than 3 letters, hence amplitude of saccades become longer[38,39].

In this study, we aimed to identify the key features of the text that allow characterising saccadic movements and can be included in the proposed model. To this end we examined the distances between syllables and words in the text. Specifically, we analysed the distances between first and second neighbours, distances between words, and distances between words that have more than 3 letters. The experimental values obtained for these distances did not exceed 3 units. Based on these findings, we divided the interval $(0,3)$ into 20 equal parts to create a discrete domain $\{d_1, d_2, \ldots, d_{20}\}$ for this variable. The normalised frequency polygons for each of these distances are presented in Fig. 2, illustrating the distribution of each type of distance found across the text. The discretized values of d_i in which we group the distances considered are represented on the horizontal axis and the vertical axis indicates their normalised frequency.

To model the probability of the saccade amplitude we propose a linear combination of the frequencies of the described distances. In order to allow for regressions (backward saccades) we include in the linear combination the four distances considered (i.e. first and second neighbours distances, distances between words and distances between words with more than 3 letters), but defined for the negative domain $\{d_{-20} = -d_{20}, \ldots, d_{-2} = -d_2, d_{-1} = -d_1\}$. This is, we use a negative sign to indicate the amplitude of a backward saccade. In this way, the saccade amplitudes take values between -3 and 3 and are governed by a PDF that is a linear combination of the frequencies of appearance of the different distances that characterise the text.

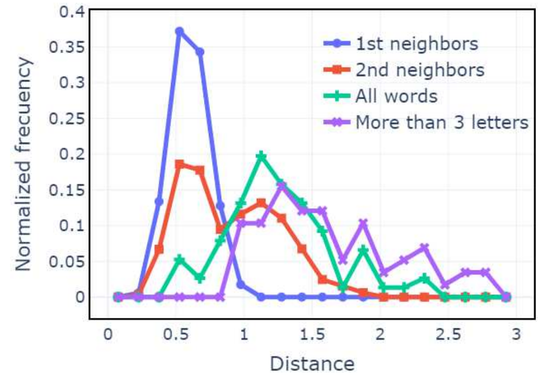


FIG. 2. Normalised frequency polygons of the distribution of first and second neighbours distances, as well as distances between words and words with more than three letters associated with the values found in the text used in the present experiment.

To give a mathematical formalism, we propose 8 functions $h_i, i = 1, 2, \ldots, 8$, 4 of them used to model the frequency of forward (positive) displacements and the other 4 used to model the frequency of backward (negative) displacements. The probability p_j of a saccade of amplitude d_j can be expressed as a fraction, with the numerator being the sum of the product of each of the eight functions h_i evaluated at d_j and their corresponding weights w_i, and the denominator being the sum of the same expression over all possible distances in the domain $[-20, 20]$.

$$p_j = \frac{\sum_{i=1}^{8} h_i(d_j)\, w_i}{\sum_{j=-20}^{20} \sum_{i=1}^{8} h_i(d_j)\, w_i}, \tag{2}$$

where $w_i, i = 1, \ldots, 8$ are the coefficients (weights) corresponding to each h_i in the linear combination. They were estimated for dyslexic and TD children from the experimental data using the curve_fit method of the scipy.optimize python package and the values found are listed in Table I. The resulting frequencies of the distances for the dyslexic children and the corresponding curve fit are plotted in Fig. 3.

The values w_i reflect the influence of each type of distance considered in the description of saccadic movements. The results presented in Table I show that children with dyslexia mainly perform saccadic movements with amplitudes that match those of the first neighbours ($w_i = 0.41$, the highest value of w_i) and, to a lesser extent, the amplitude of their saccades correspond to the distance of the second neighbours ($w_i = 0.2$). Backward saccades are mostly directed to first neighbours ($w_i = 0.11$). In contrast, TD children perform a

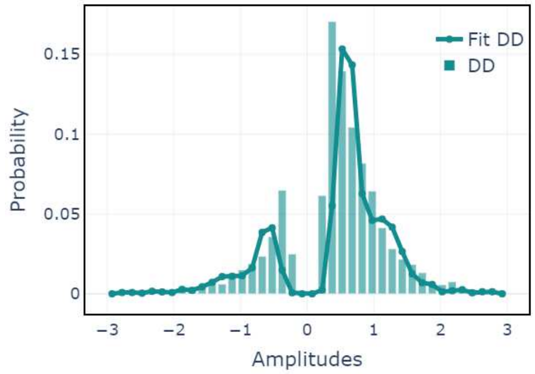


FIG. 3. Histogram of the amplitude of saccades of dyslexic children fitted with a linear combination of functions h_i describing the frequency of occurrence of the four distances with which the text was characterised.

TABLE I. Weights w_i of the functions h_i obtained from fitting the histograms of distances between syllables and words to the empirical values of saccade amplitudes of dyslexic and TD children (Eq. (2)).

Distances (h_i)	Weight (w_i)	
	TD	DD
First neighbour backward	[0.13 ± 0.03]	[0.11 ± 0.07]
Second neighbour backward	[0.05 ± 0.09]	[0.0 ± 0.2]
All words backward	[0.0 ± 0.2]	[0.0 ± 0.4]
Words with more than 3 letters backward	[0.06 ± 0.08]	[0.0 ± 0.2]
First neighbour forward	[0.26 ± 0.03]	[0.41 ± 0.07]
Second neighbour forward	[0.07 ± 0.09]	[0.2 ± 0.2]
All words forward	[0.0 ± 0.2]	[0.0 ± 0.4]
Words with more than 3 letters forward	[0.27 ± 0.08]	[0.0 ± 0.2]

combination of jumps with amplitudes that match first neighbours distances and distances between words with a minimum of 3 letters ($w_i = 26$ and $w_i = 27$ respectively). Their backwards saccades also correspond to jumps between syllables ($w_i = 0.13$).

C. Duration of saccades

The duration of saccades can be estimated using the model proposed by Frapiccini et al[20], which has been proven to be efficient for the description of various processes, particularly reading. This model allows an accurate representation of the saccadic eye movements with a single free parameter: the amplitude. Specifically, the duration of saccades can be expressed as:

$$\tau_i^{sac} = a\ \eta_i{}^{2/5} \tag{3}$$

where η_i is the amplitude of the i-th saccadic movement and for the value of a we have obtained $a = [0.0745 \pm 0.0006]$.

IV. MODEL OF EYE MOVEMENTS DURING READING AS A CTRW

Eye movements during reading consist of moments of stillness (fixations) followed by forward or backward movements of variable lengths. The stochastic characteristics of the duration of fixations and the amplitude and duration of saccades convert the process into one of the CTRW type.

There are various approaches to the CTRW. Some of them are based on the Master equations[13,14]. Some others are derived from Langevin equations[18,40]. In this section we propose a model of eye movements during reading based on this last approach.

According to previous proposals[10,20], the equation describing the dynamics of a series of alternating jumps and waiting times is given by:

$$\frac{dx}{dt} = F(t) = \sum_i f_i(t, t_i), \tag{4}$$

where x is the horizontal coordinate of the gaze position and $F(t)$ is the eye movement activation function, represented by the sum of terms $f_i(t,t_i)$[20] with t_i the instant in which the i-th jump occurs.

The function $f_i(t,t_i)$ can be defined by the amplitude of the saccade η_i and the fixation time τ_i^{fix}. These two parameters have stochastic characteristics and comply with a joint probability density function $\Psi(\eta, \tau)$. We base our proposal on the hypothesis that the duration of fixations and the amplitude of saccades are independent, so $\Psi(\eta, \tau)$ can be written as the product of two probability functions[13]:

$$\Psi(\eta, \tau) = \lambda(\eta)\ \psi(\tau), \tag{5}$$

where $\lambda(\eta)$ and $\psi(\tau)$ are the PDFs that define the amplitude of the saccadic movement and the duration of fixation respectively. Based on the description of saccade amplitudes proposed in Sec. III B, in our model of reading the PDF $\lambda(\eta)$ is approximated by the fit of the probabilities p_j given in Eq. 2, this is the curve plotted in Fig. 3. The PDF describing the duration of fixations is given by the lognormal fit plotted in Fig. 1

In order to move the eyes a distance η, a time necessarily intervenes, this is the time of flight of the saccadic movement. After each jump there is a waiting time corresponding to the fixation duration. The sum of these two times (time of flight + fixation time) defines the intersaccadic time. In the present proposal, we do not focus on the shape of the saccadic movement but we account for the time required to move the eyes from one point to the next (time of flight). This is, we approach the process as a sequence of sudden horizontal movements that can be modelled as $x_i\delta(t - t_i)$, assuming that the jumps occur at times t_i and are followed by an intersaccadic time τ_i given by the expression:

$$\tau_i = \tau_i^{fix} + \tau_i^{sac}, \tag{6}$$

with τ_i^{fix} the duration of the i-th fixation.

Alternatively to Eq. (4), a set of subordinate equations can be considered. This approach has been discussed by Fogedby[18] and used to analyse Lévy flights. A system of subordinate equations involves an equation for the time and another for the jumps. Kleinhans[12], based on the work of Fogedby[18], discussed a discrete version of subordinated equations that provides an alternative form of the CTRW. Following Kleinhans work[12], and adapting it to eye movement signals, we associate each jump with a saccade and each waiting time with an intersaccadic time. Our approach to reading dynamics, alternative to the one governed by Eq. (4), is described by the following discrete iterative equations[16,41]:

$$x_i = x_{i-1} + \eta_i \tag{7a}$$

$$t_i = t_{i-1} + \tau_i \tag{7b}$$

were η_i and τ_i are stochastic variables derived from the PDF $\Psi(\eta, \tau)$ defined in Eq. (5). Even when the PDFs are assumed independent, we still have a coupling due to the presence of τ_i which depends on η_i through Eq. (3). Thus the expression for the waiting time is

$$t_i = t_{i-1} + \tau_i^{fix} + a\ \eta_i^{2/5}. \tag{8}$$

The probability of η_i is given by the linear combination in Eq. (2), and the PDF for τ_i^{fix} is given by Eq. (1).

Essentially, Eq. (7a) defines a set of points (positions on the text) occurring at time intervals determined by intersaccadic times which follow a stochastic dynamics described by Eq. (7b). To recreate a continuous time process, it is necessary to consider the position of the gaze remaining still until the new movement takes place[12]. This is:

$$x(t) = x_i \quad for \quad t_i \leq t \leq t_{i+1}. \tag{9}$$

To generate a synthetic signal simulating eye movements during reading, we propose an iterative algorithm governed by Eq. (7a) and Eq. (7b), requiring as input the probability of saccadic amplitudes (Eq. 2) and duration of fixations (lognormal fit plotted in Fig. 1). In each step the algorithm randomly selects a saccade amplitude and it generates a position x_i on the text from the sum of the previous position and the selected amplitude. Independently, randomly selects a fixation duration and generates time increments t_i from the sum of the previous time and the selected fixation duration, and the duration of the saccade selected.

1. To simulate each line in text:
   (a) Initialise the gaze position x_0 in the position of the first syllable of the line.
   (b) Obtain a realisation for η_i using the probability defined in Eq. (2).
   (c) Check position: if $x_i + \eta_i$ is greater than the position of the first syllable go to the next step. if not, η_i is discarded and a new realisation is requested (go to step 1a). This is so that the synthetic signal does not have a movement to the left where there is no text to read.
   (d) Check end of line: if $x_i + \eta_i$ is smaller than the position of the last syllable of the line go to the next step. If not, this value of η_i is discarded and pass to the next line. For the new line repeat from the step 1. If it was already in the last line, the step 1 is finished.
   (e) If 1c and 1d are checked, the new position is $x_{i+1} = x_i + \eta_i$. Then go to step 1b.
2. Calculate the duration of each saccade (flight time) τ_i^{sac}.
3. Using the lognormal PDF (Eq.(1)) generate as many duration of fixations τ_i^{fix} as saccades were generated.
4. Define $t_{i+1} = t_i + \tau_i^{fix} + \tau_i^{sac}$.

As a result we obtain a time series $\{(t_0 = 0, x_0), (t_1, x_1), \ldots (t_N, x_N)\}$ indicating the position x_i of the eyes on the text at each time t_i in which a saccade occurs. For the intermediate instants $t \in (t_i, t_{i+1})$ we consider the eyes remain still (see Eq. (9)), thus eye movements are represented as a step function.

We base our proposal on the assumption that the probability of a saccade amplitude is independent of the probability of the following duration of fixation. This does not necessary implies lack of memory in the process. As most cognitive processes, there is a persistence on the process due to the fact that the subject wants to perform the task assigned. This persistence is included through the probability of the amplitude of saccades given by Eq. (2) and can be visualised in Fig. 3, where it is evident that forward (positive) saccades are more probable.

V. RESULTS

With the proposed algorithm we generated 50 synthetic signals simulating the eye movements of dyslexic children reading the text used in the present experiment. To evaluate the resulting signals we propose two direct inspections and a more elaborated analysis based on the entropy and complexity of actual and synthetic data.

A. Direct comparison

In order to directly compare synthetic signals with the experimental ones, we employed two different methods. The first method consisted of visually inspecting trajectories of the synthetic and actual signals having approximately the same duration, by plotting them together and comparing their shapes. An example of this plot is shown in Figure 4. The plot shows the x-coordinates (horizontal eye movements) of one empirical and one synthetic signal generated by our stochastic model. Our visual inspections showed that the synthetic signals closely matched the actual signals in terms of reading speed and trajectory shape. In fact, in some cases, the signals were so similar that they almost overlapped each other.

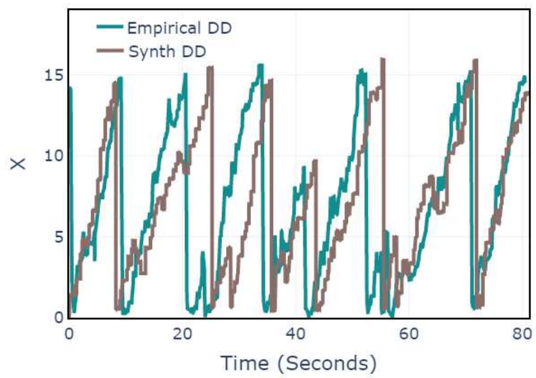


FIG. 4. Visual comparison of empirical and modelled horizontal eye movement signals during reading a text. The model closely reproduces the dynamics of the measured signal, validating its ability to simulate the complexities of reading.

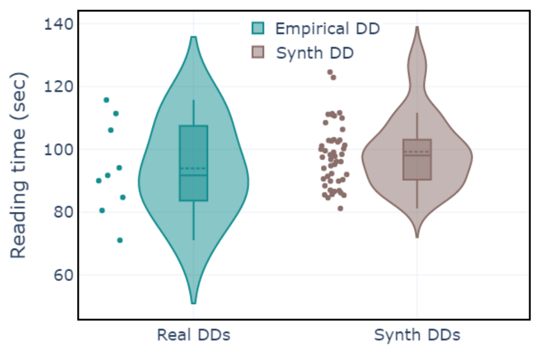


FIG. 5. Reading times for actual data and synthetic ones.

The second method used to compare synthetic versus actual signals was by comparing the resulting reading times. In this case, all signals can be directly compared, as it can be observed in Fig. 5 where the boxplots summarise the reading time of the 12 dyslexic readers and that of the simulated data.

The agreement found between synthetic and empirical signals indicate that our proposal provides a simple and effective method to globally model eye movements of dyslexic children during reading.

B. Complexity-entropy profiles

Shannon entropy and statistical complexity are two properties that provide insights about the randomness or hidden structures of a given feature[31]. These indices have been previously used in the study of eye tracking signals related to a cognitive process[25].

Here we analyse the entropy and complexity properties of both actual and synthetic data in order to compare two features: the amplitude and direction of saccades on the one hand, and the organisation of fixation locations on the text,

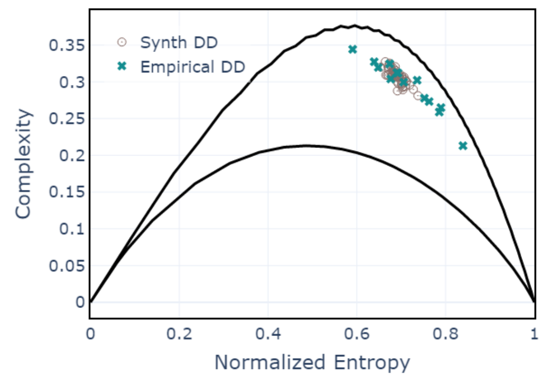


FIG. 6. Shannon entropy and statistical complexity associated to the amplitude of saccades of actual and synthetic data.

on the other. The methodology used is the one proposed by Rosso and collaborators[31].

The Shannon entropy and the statistical complexity associated to the histogram of the amplitude of saccades were calculated for each empirical and synthetic signal. We kept the previous considerations of using negative amplitudes to indicate regressions, and took the same 40 bins in the interval $(-3,3)$ (see Sec.III B). The values obtained are plotted in the Complexity-Entropy (C-H) plane in Fig. 6. As it was expected, empirical and synthetic signals locate in the same region of the plane, indicating that the spatial distribution of the data generated with the model and their statistical properties coincide with those of actual datasets. Given the location of the data in the C-H plane, it can be said that the configuration of amplitude and direction of saccades tends to behave unpredictably (entropy higher than 0.6) and reaches the maximum complexity.

Another key element in the structure of the signals is the ordinal relation between fixation locations. This is, the organisation of the sequence of position of fixations on the text. To compare this structure we performed a complexity-entropy analysis of fixation locations using the permutation distribution, originally proposed by Bandt & Pompe[42]. The methodology consists on taking sliding windows of fixed length over the time series $\{x_t\}$, each of which is associated with a permutation pattern according to the ordinal position of the data points in it. The relative frequency of these patterns defines the probability distribution. The recommended number q of permutation patterns is considered to be less than $\frac{1}{5}$ of the actual duration of the time series to ensure meaningful sampling. In this work, we set a window size of 4 consecutive points ($ws = 4$) which leads to $4! = 24$ possible permutation patterns, below the recommended maximum. This distribution captures the spatial organisation of horizontal movements of the eyes.

The values of Shannon entropy and statistical complexity obtained for actual and synthetic data are plotted in Fig. 7. As in the previous comparisons, we found a very good agreement between both sets of data, which reinforces our conclusion that the model was able to reproduce the eye movement dy-

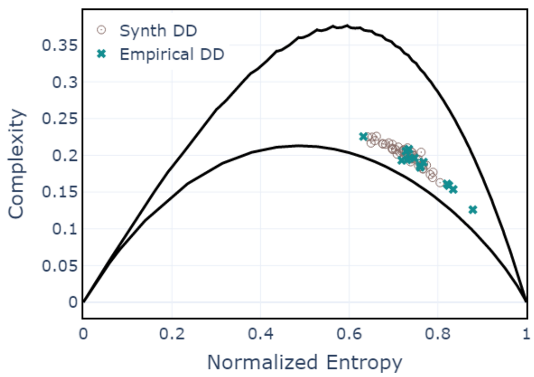


FIG. 7. Entropy and complexity of actual and synthetic signals calculated with the ordinal patterns distribution.

namics of dyslexic children reading the experimental text.

VI. DISCUSSION AND CONCLUSION

The study of the dynamics of eye movements in cognitive tasks is considered of great importance due to the link existing between eye movements and the cognitive process itself. In this proposal we have presented some contributions on this subject. We have addressed the modelling of the eye movements that dyslexic children perform while reading a text in Spanish. The structure of these movements can be considered the most elementary within the range of different structures emerging during reading as the reader becomes an expert. The eye movement dynamics of children with dyslexia closely resemble the eye movements of children in the early stages of learning to read.

This work has focused on three main objectives. One was the analysis of eye movements recorded from dyslexic children during reading a text in Spanish. The second purpose was the introduction of a simple but efficient model that replicates the studied movements. A third objective was to analyse the capability of the model to reproduce certain structures that characterise the eye movement dynamics under study.

To achieve these objectives we analysed, on the experimental data, four fundamental features that govern the eye movement dynamics during reading: the duration of fixations and the amplitude, direction and duration of saccades. Particularly, we provided expressions for the probability of these events. We have shown that the PDF that characterise the duration of fixations can be modelled with a lognormal function, and exhibited that a linear combination of numerical functions, obtained from distances between syllables and words in the text, can be used to describe the probability associated to the amplitude and direction of saccades. Duration of saccades were obtained from their amplitude, as proposed in a previous work[20]. All these elements allowed us to propose a stochastic model for the dynamics of eye movements of dyslexic children during reading, based on the formalism of the CTRW. The movements were decomposed into instantaneous jumps and waiting times by means of a system of two coupled finite difference equations. An analysis of certain patterns composing the sequence of fixations and saccades was carried out in terms of statistical complexity and entropy which allowed, at the same time, revealing the behaviour of the studied structures and comparing actual eye movement signals with synthetic ones generated with the model.

The proposed model can be considered as a continuation of the ideas previously introduced in the works of Specht et al.[10], Bouzat et al.[19] and Frapiccini et al.[20]. Our proposal combines the model of saccadic movements with stochastic variables as the amplitude and direction of saccades and the duration of fixations. Such a combination gives rise to a CTRW model based on a Langevin equation. Instead of directly solving the Langevin equation, we decomposed it into a system of two subordinated finite difference equations similar to those used by Kleinhans[12].

The analysis performed on synthetic data generated with an algorithm based on the proposed model, indicates that the simulated eye movements behave similar to the actual ones. The visual inspection and the comparison of reading time gave very good agreements. This type of comparison is not conclusive, which is why we implemented the complexity-entropy analysis, a methodology that allows studying complex structures.

The results obtained for the amplitude and direction of saccades showed a very good agreement between actual and synthetic data, at the same time that it revealed a configuration of amplitude and direction of saccades tending to uncertainty with a maximum complexity. It can be interpreted as an intermediate state between a predictable order and an unpredictable disorder. The coincidences obtained for both empirical and synthetic data in this analysis was actually not surprising as the probability distributions used to calculate the entropy and complexity indices were all obtained form the same one. The important conclusion from this result is that the linear combination proposed to model the amplitude and direction of saccades, which was based on the distances between syllables and words in the text, provides a very good description of actual saccades. As was noticed from the analysis of Fig. 6, synthetic signals are concentrated in the middle of the point cloud determined by the actual data. This can be explained by the fact that actual data present more variability than the synthetic ones. A term of error (residue) can be included in the linear combination in Eq. (2) to obtain a better approach to the variability of actual eye movement signals.

A second complexity-entropy analysis was carried out in order to study the configuration of position of fixations on the text. We analysed the ordinal relation between them using ordinal patterns as proposed by Bandt & Pompe[42]. The agreement found reveal that the model is capable of reproducing the dynamics of patterns of fixation locations, an element that is not included in the generating algorithm. In this case, the inclusion of an error term in Eq. (2) seems not to be necessary.

One final consideration can be made with respect to our proposed model. For a different text, the functions h_i in Eq. (2) must be estimated again since they are based on the distances characterising the syllables and words on it. We hypothesise that the coefficients w_i will remain similar, based on the fact that the behaviour of dyslexic children will not change. Thus, if the hypothesis is correct, the model allows generating synthetic signals from the analysis of a few records.

ACKNOWLEDGMENTS

We are grateful to Mg. Liliana Fonseca and collaborators for the assistance in matters of experimental data recording and diagnosis and to the insightful comments of the professionals at the Centro Integral de Neurociencias Aplicadas of Bahia Blanca. This research was developed as a part of the projects UNS PGI 24/F078, CONICET PIP KE3 11220200102879CO and by Agencia Nacional de Promocion Cientifica y Tecnológica PICT - 2020 - SERIE A - 02450.

DATA AVAILABILITY

The data that supports the findings of this study is available from the corresponding author upon request.

REFERENCES

[1] J. Hyönä and R. K. Olson, "Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency," J. Exp. Psychol. Learn. Mem. Cogn. **21**, 1430–1440 (1995).

[2] F. Hutzler and H. Wimmer, "Eye movements of dyslexic children when reading in a regular orthography," Brain Lang. **89**, 235–242 (2004).

[3] C. Prado, M. Dubois, and S. Valdois, "The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span," Vision Res. **47**, 2521–2530 (2007).

[4] K. Rayner, "Eye movements in reading: models and data," J. Eye Mov. Res. **2**, 1–10 (2009).

[5] R. Engbert, A. Nuthmann, E. M. Richter, and R. Kliegl, "Swift: A dynamical model of saccade generation during reading," Psychol. Rev. **112**, 777–813 (2005).

[6] H. Jarodzka and S. Brand-Gruwel, "The neurologycal basis of developmental dyslexia. an overview and working hypothesis," J. Comput. Assist. Learn. **33**, 193–201 (2017).

[7] F. J. Martos and J. Vila, "Differences in eye movements control among dyslexic, retarded and normal readers in the spanish population," Read Writ **2**, 175–188 (1990).

[8] G. F. Eden, J. F. Stein, H. M. Wood, and F. B. Wood, "Differences in eye movements and reading problems in dyslexic and normal children," Vision Res. **34**, 1345–1358 (1994).

[9] M. N. B. andG. Öqvist Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for dyslexia using eye tracking during reading," PLoS ONE **11**, e0165508 (2016).

[10] J. I. Specht, L. Dimieri, E. Urdapilleta, and G. Gasaneo, "Minimal dynamical description of eye movements," Eur. Phys. J. B **90**, 1–12 (2017).

[11] J. Pekkanen and O. Lappi, "A new and general approach to signal denoising and eye movement classification based on segmented linear regression," Sci. Rep. **7**, 17726 (2017).

[12] D. Kleinhans and R. Friedrich, "Continuous-time random walks: Simulation of continuous trajectories," Physical Review E **76**, 061102 (2007).

[13] E. Scalas, R. Gorenflo, and F. Mainardi, "Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation," Phys. Rev. E **69**, 011107 (2004).

[14] I. Gomez Portillo, D. Campos, and V. Méndez, "Intermittent random walks: transport regimes and implications on search strategies," Journal of Statistical Mechanics: Theory and Experiment **2011**, P02033 (2011).

[15] G. Boccignone, "Advanced statistical methods for eye movement analysis and modelling: A gentle introduction." (Springer, Cham., 2019) Chap. 9, pp. 309–405.

[16] R. Metzler and J. Klafter, "The random walk's guide to anomalous diffusion: a fractional dynamics approach," Physics reports **339**, 1–77 (2000).

[17] H. G. Othmer, S. R. Dunbar, and W. Alt, "Models of dispersal in biological systems," Journal of mathematical biology **26**, 263–298 (1988).

[18] H. C. Fogedby, "Langevin equations for continuous time lévy flights," Physical Review E **50**, 1657 (1994).

[19] S. Bouzat, F. M. L. Freije, and G. Gasaneo, "Inertial movements of the iris as the origin of postsaccadic oscillations," Phys. Rev. Lett. **120**, 178101 (2018).

[20] A. L. Frapiccini, J. A. Del Punta, K. V. Rodriguez, L. Dimieri, and G. Gasaneo, "A simple model to analyse the activation force in eyeball movements," The European Physical Journal B **93**, 1–10 (2020).

[21] D. G. Stephen and D. Mirman, "Interactions dominate the dynamics of visual cognition," Cognition **115**, 154–165 (2010).

[22] D. J. Aks, G. J. Zelinsky, and J. C. Sprott, "Memory across eye-movements: 1/f dynamics in vision search," Nonlinear Dynamics Psychol. Life Sci. **6**, 1–25 (2002).

[23] D. Mirman, J. R. Irwin, and D. G. Stephen, "Eye movement dynamics and cognitive self-organization in typical and atypical development," Cogn. Neurodyn. **6**, 61–73 (2012).

[24] M. L. Freije, A. A. J. Gandica, J. I. Specht, G. Gasaneo, C. A. Delrieux, B. Stosic, T. Stosic, and R. de Luis-Garcia, "Multifractal detrended fluctuation analysis of eye-tracking data," in *VipIMAGE 2017. ECCOMAS 2017. Lecture Notes in Computational Vision and Biomechanics*, edited by J. Tavares and R. N. Jorge (International Publishing AG, 2017) pp. 476–484.

[25] F. R. Iaconis, A. A. J. Gandica, J. A. D. Punta, C. A. Delrieux, and G. Gasaneo, "Information-theoretic characterization of eye-tracking signals with relation to cognitive tasks," Chaos **31**, 033107 (2021).

[26] J. G. Holden, G. C. V. Orden, and M. T. Turvey, "Dispersion of response times reveals cognitive dynamics," Psychol. Rev. **116**, 318–342 (2009).

[27] E. A. Ihlen and B. Vereijken, "Interaction-dominant dynamics in human cognition: Beyond $1/f^\alpha$ fluctuation," J. Exp. Psychol. **139**, 436–463 (2010).

[28] G. Feng, "Eye movements as time-series random variables: a stochastic model of eye movement control in reading," Cogn. Syst. Res. **7**, 70–95 (2006).

[29] D. R. Chialvo, "Life at the edge: complexity and criticality in biological function," Acta Physica Polonica B **49**, 1955–1979 (2018).

[30] L. T. Martín, A. Plastino, and O. A. Rosso, "Generalized statistical complexity measures: geometrical and analytical properties," Phys. A **369**, 439–462 (2006).

[31] O. A. Rosso, M. T. Matín, H. A. Larrondo, A. M. Kowalsky, and A. Plastino, "Concepts and recent advances in generalized information measures and statistics," (Bentham Science, 2013) Chap. 8, pp. 169–215.

[32] A. P. Association, *Diagnostic and Statistical manual of mental disorders* (APA, 2013).

[33] M. Habib, "The neurologycal basis of developmental dyslexia. an overview and working hypothesis," Brain **123**, 2373–2399 (2000).

[34] M. J. Snowling, "From language to reading and dyslexia 1," Dyslexia **7**, 37–46 (2001).

[35] C. R. Lyon, C. E. Shaywitz, and B. A. Shaywitz, "A definition of dyslexia," Ann. Dyslexia **53**, 1–14 (2003).

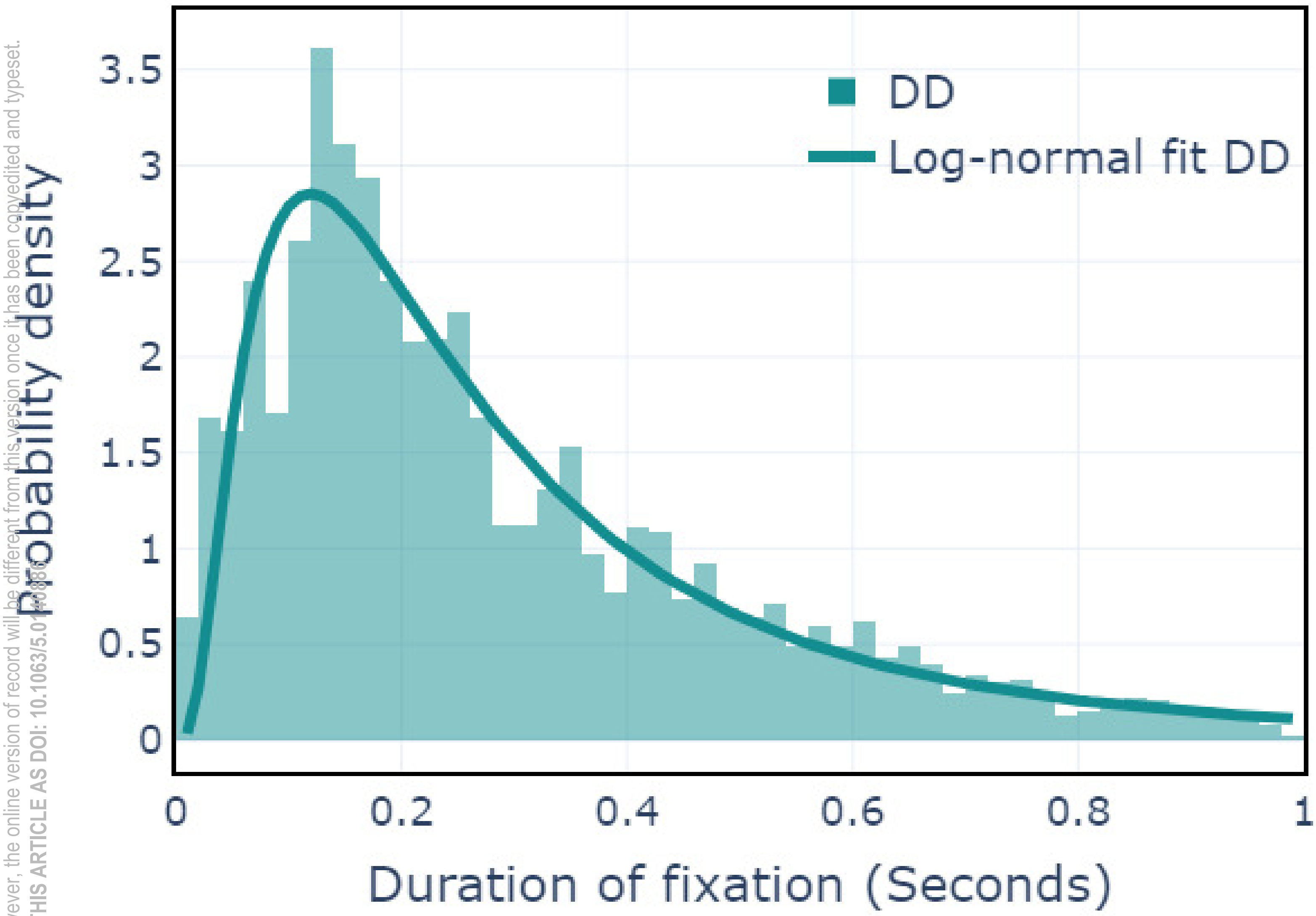
[36] M. Nyström and K. Holmkvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," Behav. Res. Methods **42**, 188–204 (2010).

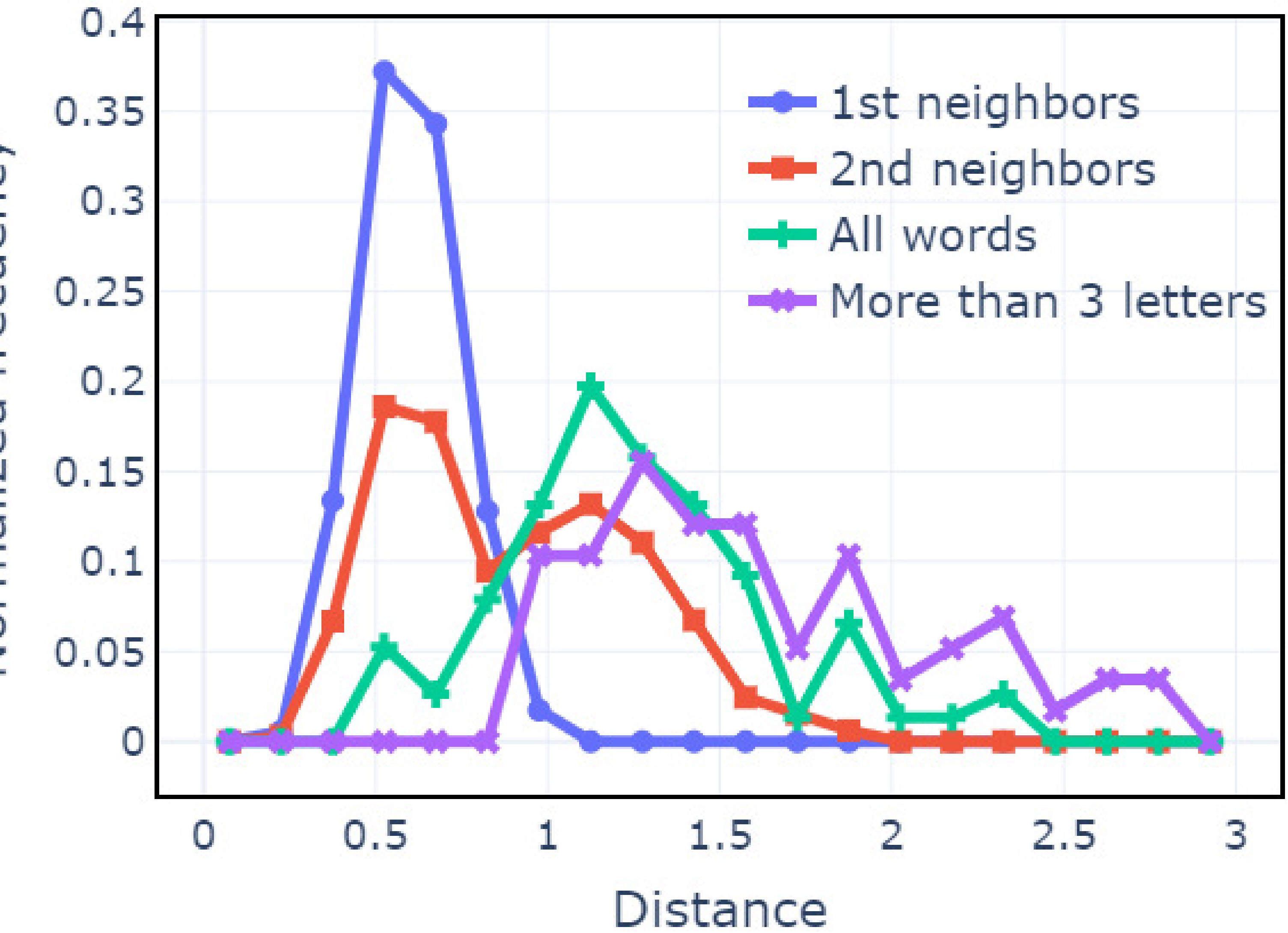
[37] S. Dehaene, "El cerebro lector: Últimas noticias de las neurociencias sobre la lectura, la enseñanza, el aprendizaje y la dislexia," in *El cerebro lector: últimas noticias de las neurociencias sobre la lectura, la enseñanza, el aprendizaje y la dislexia* (2014) pp. 444–444.

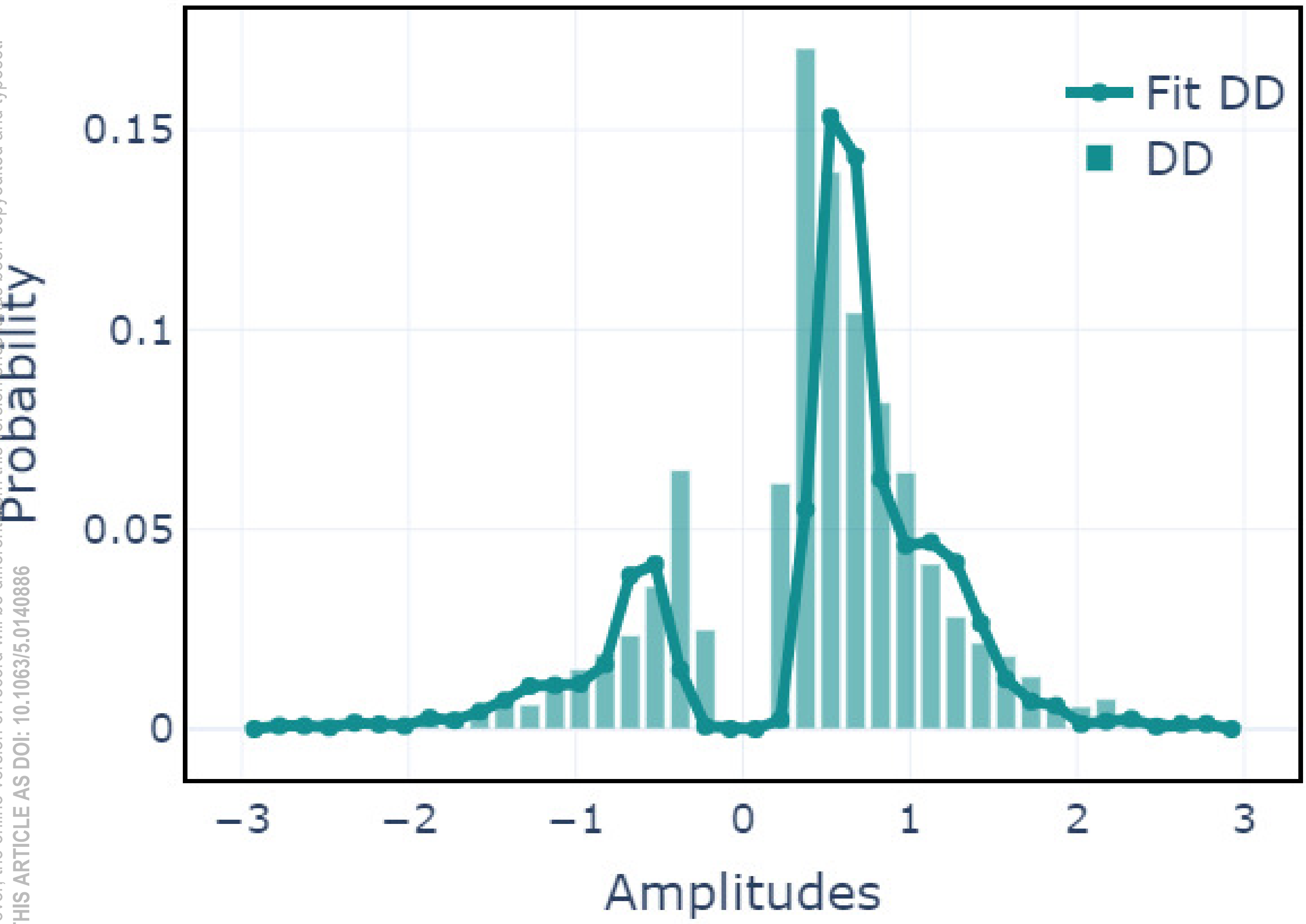
[38] M. De Luca, M. Borrelli, A. Judica, D. Spinelli, and P. Zoccolotti, "Reading words and pseudowords: An eye movement study of developmental dyslexia," Brain and language **80**, 617–626 (2002).

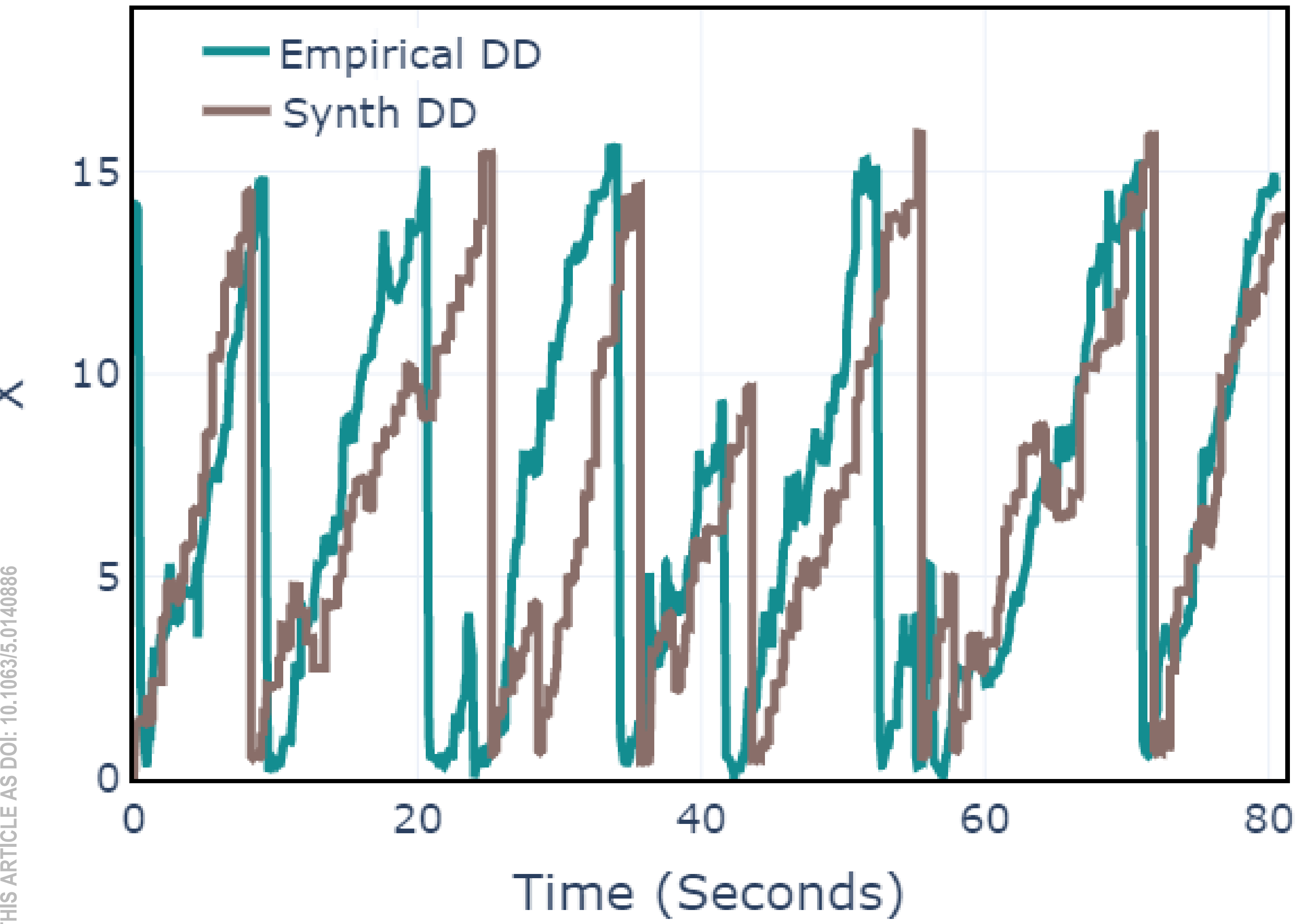
[39] M. De Luca, E. Di Pace, A. Judica, D. Spinelli, and P. Zoccolotti, "Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia," Neuropsychologia **37**, 1407–1420 (1999).

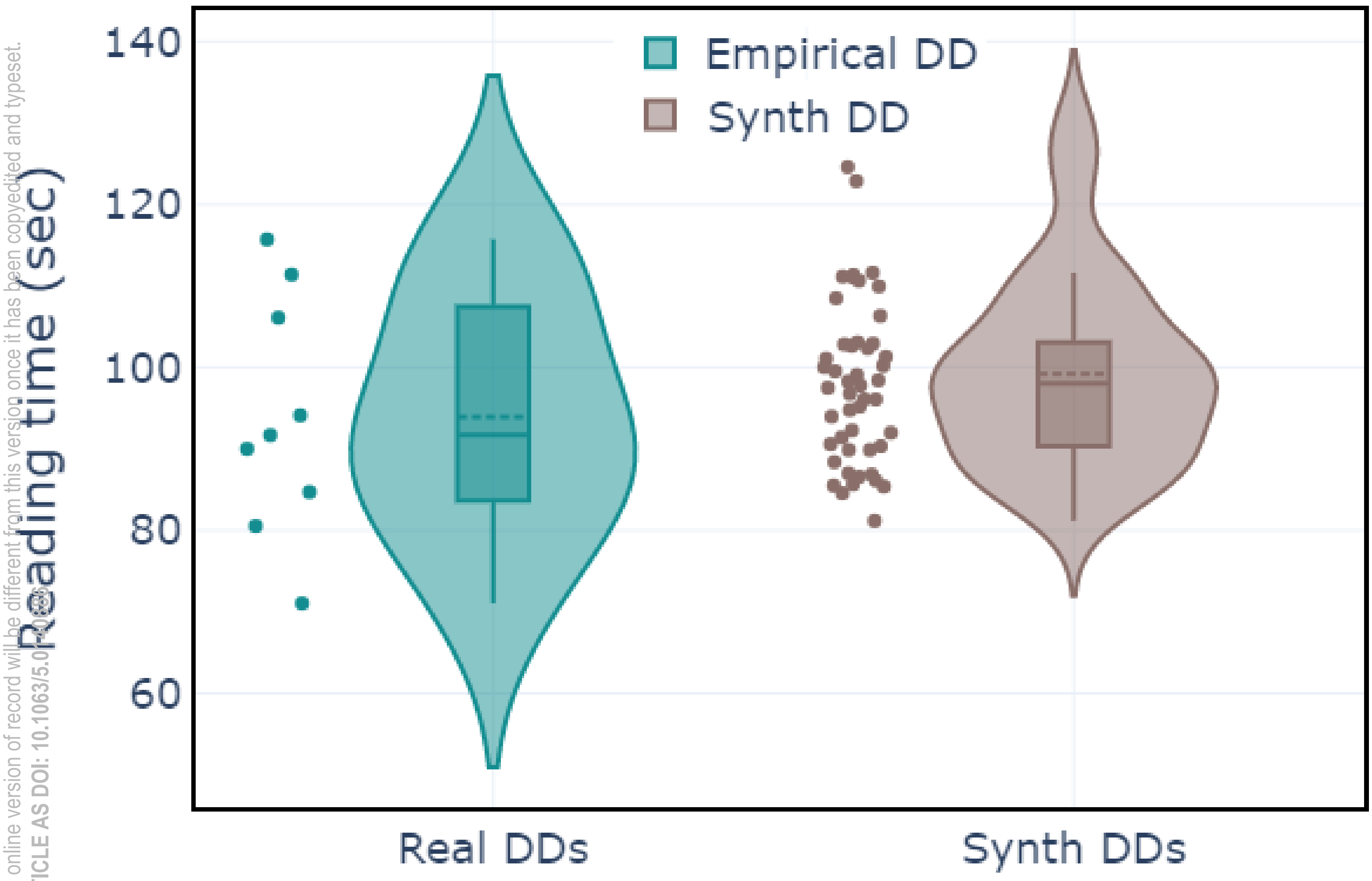
[40]S. Eule, V. Zaburdaev, R. Friedrich, and T. Geisel, "Langevin description of superdiffusive levy processes," Phys. Rev. E. **86**, 041134 (2012).
[41]G. H. Weiss and G. H. Weiss, *Aspects and applications of the random walk* (Elsevier Science & Technology, 1994).
[42]C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," Phys. Rev. Lett. **88**, 174102 (2002).

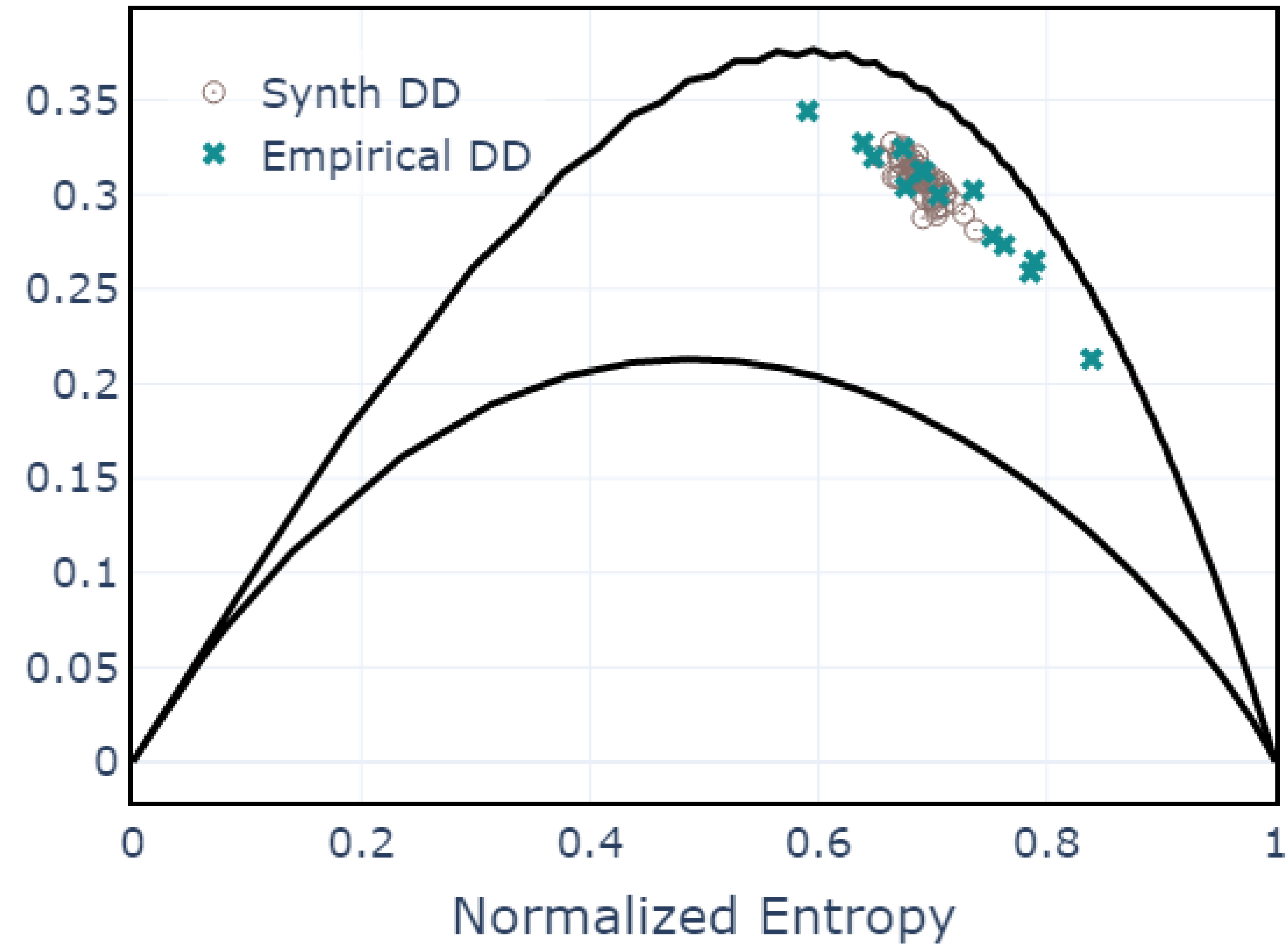

DD
Log-normal fit DD
Probability density
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Duration of fixation (Seconds)

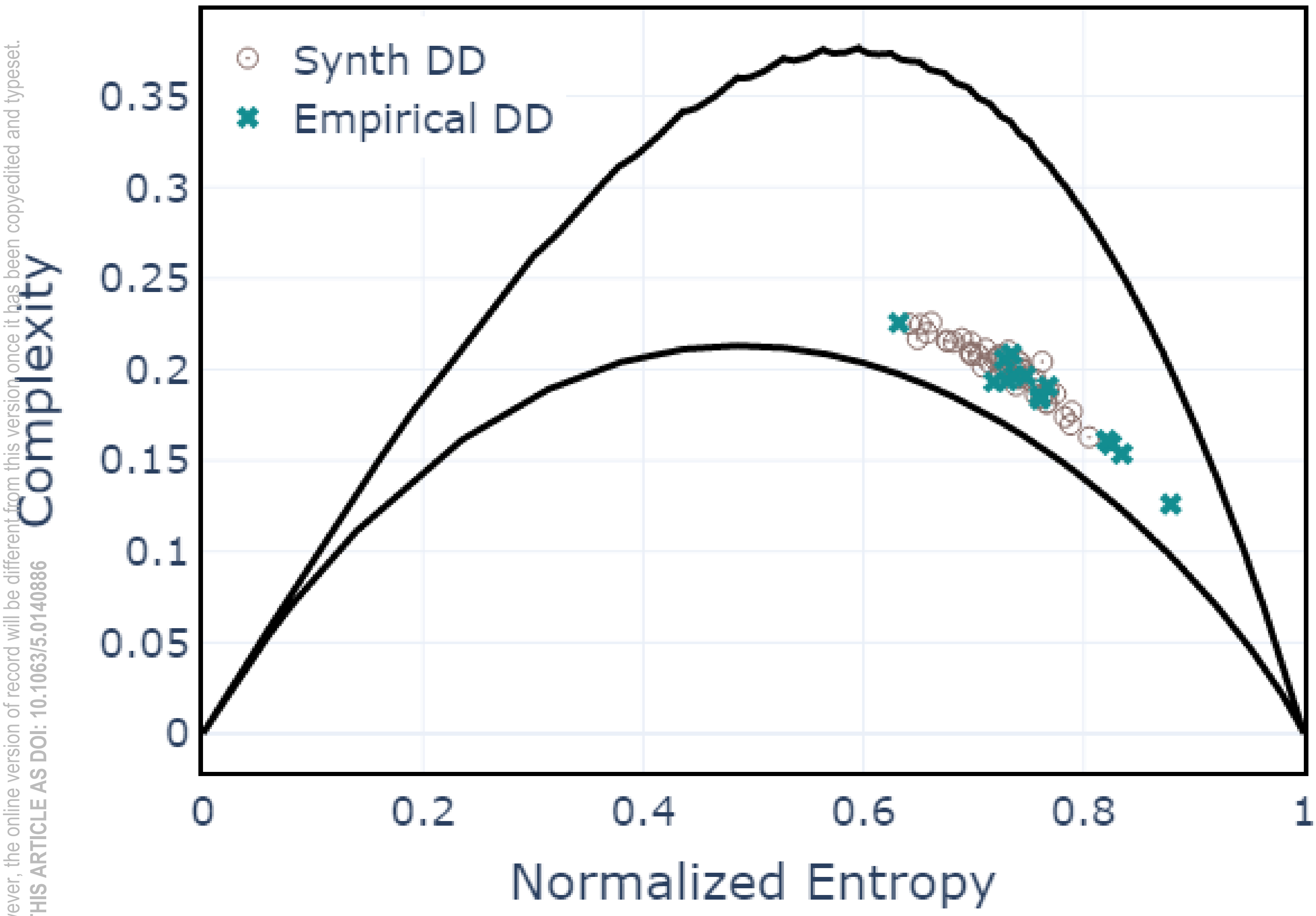

1st neighbors
2nd neighbors
All words
More than 3 letters
Normalized frecuency
Distance
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0
0.5
1
1.5
2
2.5
3


Fit DD
DD
Probability
0.15
0.1
0.05
0
−3
−2
−1
0
1
2
3
Amplitudes


Empirical DD
Synth DD
X
0
5
10
15
0
20
40
60
80
Time (Seconds)


Empirical DD
Synth DD
Reading time (sec)
140
120
100
80
60
Real DDs
Synth DDs


Synth DD
Empirical DD
Complexity
Normalized Entropy
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0
0.2
0.4
0.6
0.8
1


Synth DD
Empirical DD
Complexity
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
Normalized Entropy