

Classifying Children with Reading Difficulties from Non-Impaired Readers via Symbolic Dynamics and Complexity Analysis of MEG Resting-State Data

Stavros I. Dimitriadis¹, Panagiotis Simos², Nikolaos A. Laskaris¹, Spiros Fotopoulos³, Jack M. Fletcher⁴, Andrew C. Papanicolaou⁵

¹Artificial Intelligence and Information Analysis Laboratory, Department of Informatics, Aristotle University, Thessaloniki, 54124, Greece. Neuroinformatics Group, Department of Informatics, Aristotle University, Thessaloniki, Greece, email:

stidimitriadis@gmail.com, laskaris@aiia.csd.auth.gr

²School of Medicine, University of Crete, Crete, Greece, email: akis.simos@gmail.com

³Electronics Laboratory, Department of Physics, University of Patras, 26504 Rio, Greece, email:

spiros@physics.upatras.gr

⁴Department of Psychology, University of Houston, Houston, Texas, TX, USA, email: JackFletcher@uh.edu

⁵Department of Pediatrics, University of Tennessee Health Science Center, and Neuroscience Institute, Le Bonheur Children's Hospital, Memphis, TN, USA, email: apapanic@uthsc.edu

Abstract— Magnetoencephalography (MEG) is a brain imaging method affording real-time temporal, and adequate spatial resolution to reveal aberrant neurophysiological function associated with dyslexia. In this study we analyzed sensor-level resting-state neuromagnetic recordings from 25 reading-disabled children and 27 non-impaired readers under the notion of symbolic dynamics and complexity analysis. We compared two techniques for estimating the complexity of MEG time-series in each of 8 frequency bands based on symbolic dynamics: (a) Lempel-Ziv complexity (LZC) entailing binarization of each MEG time series using the mean amplitude as a threshold, and (b) An approach based on the neural-gas algorithm (NG) which has been used by our group in the context of various symbolization schemes. The NG approach transforms each MEG time series to more than two symbols by learning the reconstructed manifold of each time series with a small error. Using this algorithm we computed a complexity index (CI) based on the distribution of words up to a predetermined length.

The relative performance of the two complexity indexes was assessed using a classification procedure based on k-NN and Support Vector Machines. Results revealed the capacity of CI to discriminate impaired from non-impaired readers with 80% accuracy. Corresponding performance of LZC values did not exceed 55%. These findings indicate that symbolization of MEG recordings with an appropriate neuroinformatic approach, such as the proposed CI metric, may be of value in understanding the neural dynamics of dyslexia.

Keywords: complexity; MEG; dyslexia; Symbolization; Lempel-Ziv; Symbolic Dynamics

I. INTRODUCTION (HEADING 1)

Reading involves the coordinated activity of distinct brain sub-networks, mainly in the left hemisphere, each apparently responsible for specific component reading operations [1,2].

Moreover, aberrant task-related activation profiles have been documented using functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) as potential markers of deficient brain organization in developmental reading disability (dyslexia). Resting-state fMRI data have further established strong associations between reading networks and functional connectivity in both adults and children [4-7]. MEG studies have revealed an abnormal temporal correlation of network metric time-series extracted from recordings over left temporo-parietal brain areas linked to reading skills [8] and a greater repertoire and temporal variability of dominant cross-frequency coupling over right temporo-parietal brain areas [9].

To our knowledge the present is the first study that attempted to assess LZ complexity in dyslexia. Previous investigations explored the benefits of LZ complexity over symbolic sequences derived from MEG resting-state data in traumatic brain injury [12], Alzheimer's disease [13] and epilepsy [14]. However, none of the aforementioned studies examined the potential limitations of LZ complexity in discriminating patients from age-matched controls. Here, we demonstrate the limitations of LZ complexity to discriminate reading-disabled children from non-impaired readers and we proposed a novel one that can capture the non-linear dynamics of MEG recordings.

Specifically, we compared two alternative methods for converting sensor-level MEG time-series into sets of symbols under the notion of symbolic dynamics and complexity analysis. Whereas the LZ method entails binarization of each MEG time series with mean amplitude as the appropriate threshold [15], the Complexity Index (CI) utilizes the NG algorithm [16] to transform each MEG time series to a symbolic sequence with more than two symbols [17].

Our main goal was to assess the relative classification accuracy of each metric and to support of their potential value for the study of aberrant neurophysiological activity characteristic of various developmental brain disorders.

II. METHODS

A. Participants

MEG recordings were obtained from 25 reading-disabled children (aged 12.20 ± 2.1 years) and 27 non-impaired readers (aged 11.35 ± 2.8). All subjects provided informed consent and all procedures were approved by the University of Texas-Health Science Center Institutional Review Board. Resting state MEG activity was recorded using a 248-channel Magnes WH3600 system (4D Neuroimaging Inc., San Diego, CA) at a sampling rate of 1017.25Hz for 3 minutes with eyes closed. Axial gradiometer recordings were transformed to planar gradiometer field approximations using the *sincos* method of Fieldtrip [18].

B. Elimination of Non-cerebral activity

MEG recordings were filtered offline using a third order two-pass Butterworth filter between 0.5–80Hz and a notch filter at 60Hz. The extended Infomax algorithm [18] was used to estimate independent components (ICs) on individual channel recordings. Kurtosis, skewness, temporal course and spatial distribution of ICs were used as metrics in order to identify and remove ICs associated with ocular, muscle and cardiac artifacts. Finally, non-artifact ICs were back-projected to the original 248-channel MEG space to reconstruct the original recordings. Data were subsequently bandpassed in the following conventional frequency bands: δ (0.5–4 Hz), θ (4–8 Hz), α_1 (8–10 Hz), α_2 (10–13 Hz), β_1 (13–15 Hz), β_2 (15–19 Hz), β_3 (20–29 Hz), and γ (30–45 Hz) using a zero-phase 3rd order filter in both directions using the *filtfilt* function in matlab

C. Complexity Estimation

The bandpassed MEG signals were subsequently transformed into a finite set of symbols: [0,1] for LZC, and optimized based on the reconstructed error for NG [19–21].

1) *Lempel-Ziv Complexity*: The LZC which is described in detail in [15] is an algorithm that counts different substrings in the binarized symbolic time series $STS^{LZ}=[01110\dots]$. Here we transformed frequency-dependent oscillations into a binary time series using the mean amplitude as a threshold.

2) Neural Gas (NG) Algorithm and Complexity Index (CI): An alternative method to transform the MEG signal into symbols is to adopt a proper algorithm that can learn the manifold of a reconstructed phase space and then determining the appropriate mapping between trajectories and symbols (alphabet). Here, we reconstructed each bandpassed time series into a common reconstructed space and then applied the NG algorithm to derive a set of symbols that can describe the original signal with a reduced amount of error. For details on the procedure see [21].

Each concatenated time series was first embedded in a multidimensional space as described in equation (1):

$$x(n) = [y(n); y(n+T); \dots; y(n+(d_E - 1)T)] \quad (1)$$

where the time lag T is determined using mutual information and the embedding dimension d_E is obtained using the false nearest neighbors test [22].

Having estimated the reconstructed error between the original MEG time series and the one described by the NG-derived codebook, we fixed the number of symbols for each time series. Setting the threshold for the reconstructed error rate at $< 8\%$, we found that $k=6$ symbols could adequately describe the original neuromagnetic activity. Finally, each MEG sensor was transformed to a Symbolic Time Series $STS^{NG}=[1\ 2\ 3\ 4\ 5\ 6\ 2\ 1\ \dots]$ with $k=6$.

The CI metric was computed based on the quantification of distinct words over the STS^{NG} up to a length $l=7$. CI values were then normalized using 200 randomized versions of the original symbolic sequence.

D. Classification Scheme

The discriminative capacity of each complexity index (LZC and CI) was assessed via machine learning techniques. We adopted as an appropriate feature extraction algorithm the laplacian score (LS) [23] and determined the statistical threshold applied to LS through bootstrapping. The latter was defined as the mean $+2SDs$ of 1.000 laplacian scores obtained via the randomization procedure across the entire set of features, separately for each of the complexity indexes. The original set of features consisted of 8 (frequency bands) \times 248 (MEG sensors) LZC and 8 \times 248 CI values per participant.

Next, the labels of the two participant groups were shuffled and LS was reestimated for each feature. Classification performance of LZC- and CI-related features was assessed via a k -nearest neighbor (k -NN) algorithm and Support Vector Machines (SVMs). Initially, each set of features was ranked in decreasing relative importance according to the associated LS. In successive classification runs we tested the performance of progressively increasing sets of features starting from the ones at the top of the ranked feature lists. The procedure was discontinued when classification accuracy reached a plateau.

III. RESULTS

A. Classification Performance on Complexity

Table I and II summarize the classification accuracy, sensitivity and specificity of each complexity index (LZC and CI, respectively). We adopted two classification methods (k -NN and SVMs) applied on a 10-fold cross validation scheme running over 100 iterations. Results revealed that a set of 55 CI features, originally associated with the highest LS scores, resulted in $\sim 80\%$ classification accuracy between students with dyslexia and typically achieving readers. Sensitivity and specificity estimates were in the same range. In contrast, the classification accuracy based on LZC was barely above chance level (55%; see Table II).

Table III indicates that the majority of CI features associated with best classification results were derived the θ to

β_2 frequency range whereas Figure 2 reveals a rather widely distributed pattern of LS values across the sensor array.

TABLE I. CLASSIFICATION PERFORMANCE FOR COMPLEXITY INDEX (CI) BASED ON SUBWORD LENGTH $L=7$

	Accuracy (%)	Sensitivity (%)	Specificity (%)	# of Selected features
kNN	77.32±1.12	76.15±1.25	77.03±2.12	55/(248*8)
SVM	80.36±1.14	80.19±1.41	79.28±2.17	

TABLE II. CLASSIFICATION PERFORMANCE FOR LZC

	Accuracy (%)	Sensitivity (%)	Specificity (%)	# of Selected features
kNN	54.12±1.89	53.95±1.79	54.11±2.43	39/(248*8)
SVM	55.67±1.67	56.23±1.91	55.91±2.01	

TABLE III. DISTRIBUTION OF FEATURES OVER FREQUENCIES

Band	δ	θ	α_1	α_2	β_1	β_2	β_3	γ	Total
CI	0	10	8	10	24	1	2	0	55
LZC	1	6	8	6	5	5	6	2	39

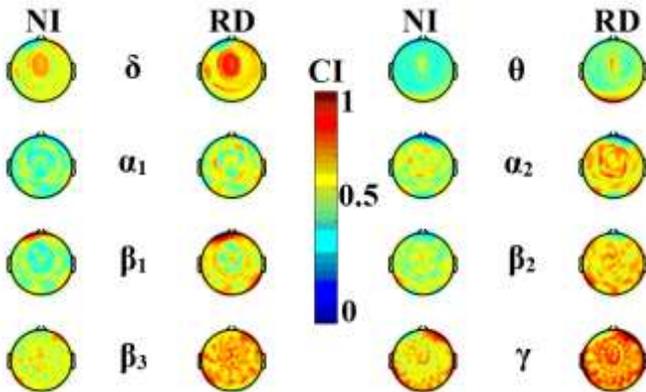


Fig.1. Topographical layouts of group-averaged CI values for each frequency band and group of participants.

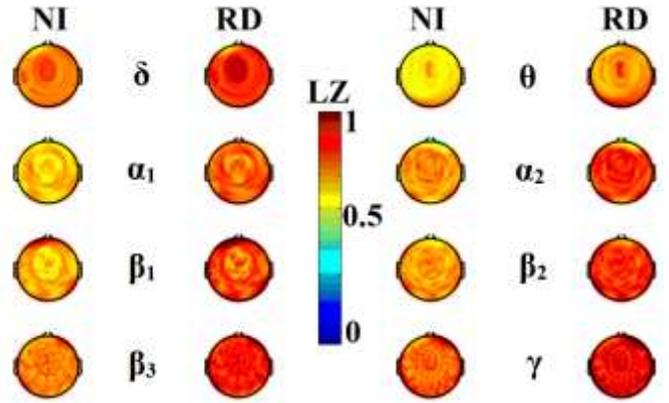


Fig.2. Topographical layouts of group-averaged LZC values for each frequency band and group of participants.

IV. DISCUSSION AND CONCLUSIONS

The present study explores how symbolic dynamics and complexity analysis of sensor-level neuromagnetic data can reveal aberrant activity in children with reading difficulties. Additionally, we demonstrated the effect of two main preprocessing steps to the final outcome: a) the appropriate neuroinformatic tool to symbolize the reconstructed space over a binarization of each MEG sensor based on the mean amplitude and b) the adaptation of a proper complexity index in both aforementioned cases.

Under the umbrella of symbolic dynamics, complexity analysis based on symbolic sequences extracted via the proposed NG scheme resulted in adequate discrimination capacity at the individual participant level. Moreover, the NG-based features clearly outperformed the LZC features. In ongoing analyses we are exploring features based on both neuromagnetic signal amplitude and phase [19] and applying symbolic transfer entropy and symbolic mutual information to construct functional brain networks [21].

ACKNOWLEDGMENT

This research was supported in part by grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health. This research was also supported by Greek State Scholarships Foundation (IKY) as part of a post-doctoral research fellowship to S.I. Dimitriadis.

REFERENCES

- [1] Norton ES, Beach SD, Gabrieli JD. "Neurobiology of dyslexia," *Curr Opin Neurobiol.* vol.30,pp.73–78,2015.
- [2] Schlaggar BL, McCandliss BD. "Development of neural systems for reading," *Ann RevNeurosci.* vol.30,pp.475–503,2007.
- [3] T.L. Richards, T.J. Grabowski, P. Boord, P. Yagle, M. Askren, Z. Mestre, *et al.* "Contrasting brain patterns of writing-related DTI parameters, fMRI connectivity, and DTI–fMRI connectivity correlations in children with

- and without dysgraphia or dyslexia," *Neuroimage Clin*, vol.8 pp. 408–421,2015.
- [4] M.S.Koyama ,A. Di Martino,C. Kelly,M.Mennes, D.R.Jutagir et al. "Resting statefunctional connectivity indexes reading competence in children and adults," *J Neurosci*, vol.31,no.23, pp.8617–8624,2011.
- [5] M.S.Koyama ,A. Di Martino,C. Kelly,D.R. Jutagir ,J. Sunshine,S.J. Schwartz , et al. "Cortical signatures of dyslexia and remediation: an intrinsic functional connectivity approach," *PLoS ONE* ,8(2):e55454,2013.
- [6] Schurz M, Wimmer H, Richlan F, Ludersdorfer P, Klackl J, Kronbichler M. "Resting state and task-based functional brain connectivity in developmental dyslexia," *Cereb Cortex*,vol.25,no.10,3502,2014.
- [7] Zhang M, Li J, Chen C, Xue G, Lu Z, Mei L, et al. "Resting-state functional connectivity and reading abilities in first and second languages," *NeuroImage*,vol.84, pp.546–553,2014.
- [8] S.I.Dimitriadis ,N.A. Laskaris ,P.G. Simos,J.M. Fletcher,R.Rezaie, A.C.Papanicolaou. " Altered temporal correlations in resting-state connectivity fluctuations in children with reading difficulties detected via MEG," *NeuroImage*,vol.83,pp307–317,2013.
- [9] S.I.Dimitriadis ,N.A. Laskaris ,P.G. Simos,J.M. Fletcher, A.C.Papanicolaou , "Greater Repertoire and Temporal Variability of Cross-Frequency Coupling (CFC) Modes in Resting-State Neuromagnetic Recordings among Children with Reading Difficulties," *Front Hum Neurosci* .Apr 26;10:163,2016.
- [10] M.S.Koyama ,C. Kelly,Z. Shehzad, D.Penesetti,F.X. Castellanos,M.P. Milham, "Reading networks at rest," *Cereb Cortex* vol.20, no.11,pp.2549–2559,2010.
- [11] M. Vourkas ,S.Micheloyannis, P.Simos, R.Rezaie , J.M.Fletcher, P.T.Cirino, A.C.Papanicolaou, "Dynamic task-specific brain network connectivity in children with severereading difficulties," *Neurosci Lett*,vol.488 no.2 , pp.123–128, 2011.
- [12] Q. Luo,D. Xu,T. Roskos,J. Stout,. Kull,X. Cheng,Diane Whitson,Erich Boomgarden,Jeffrey Gfeller,and Richard D. Bucholz, " Complexity Analysis of Resting State Magnetoencephalography Activity in Traumatic Brain Injury Patients," *J Neurotr.*, vol.30, pp.1702–1709 2013
- [13] D. Abásolo, R. Hornero, C. Gómez, M. García, and M. López,"Anal-ysis of EEG background activity in Alzheimer's disease patients withLempel-Ziv complexity and central tendency measure,"*Med. Eng.Phys.*, vol. 28, no. 4, pp. 315–322, May 2006.
- [14] N. Radhakrishnan and B. Gangadhar,"Estimating regularity in epileptic seizure time-series data,"*IEEE Eng. Med. Biol. Mag.*, vol.17, no. 3, pp. 89–94, May-Jun. 1998
- [15] A. Lempel and J. Ziv, "On the Complexity of Finite Sequences," in *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75-81, Jan 1976.
- [16] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, "Neural-gas' network for vector quantization and its application to time-series prediction," *IEEE Trans Neural Netw*, vol. 4, no. 4, pp. 558–569, 1993.
- [17] S. Janson, S. Lonardi, and W. Szpankowski, "On average sequence complexity," *Theoretical Computer Science*, vol. 326, no. 1–3, pp. 213– 227, Oct. 2004.
- [18] R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, Schoffelen,"FieldTrip: Open Source Software for Advanced Analysis of MEG,EEG, and Invasive Electrophysiological Data," *ComputationalIntelligence and Neuroscience, Computational Intelligence and Neuroscience*, vol. 2011, 2011, p. e156869, Dec. 2010
- [19] S. I. Dimitriadis, N. A. Laskaris, V. Tsirka, S. Erimaki, M. Vourkas, S. Micheloyannis, and S. Fotopoulos, "A novel symbolization scheme for multichannel recordings with emphasis on phase information and its application to differentiate EEG activity from different mental tasks," *Cogn Neurodyn*, vol. 6, no. 1, pp. 107–113, Feb. 2012.
- [20] S. I. Dimitriadis, N. A. Laskaris, and A. Tzelepi, "On the quantization of time-varying phase synchrony patterns into distinct functional connectivity microstates (FC_μstates) in a multi-trial visual ERP paradigm," *Brain Topogr*, vol. 26, no. 3, pp. 397–409, Jul. 2013.
- [21] S. Dimitriadis, Y. Sun, N. Laskaris, N. Thakor, and A. Bezerianos, "Revealing cross-frequency causal interactions during a mental arithmetic task through symbolic transfer entropy: a novel vectorquantization approach," *IEEE Trans Neural Syst Rehabil Eng*, Jan. 2016.
- [22] H.D.I. Abarbanel, *The Analysis of ObservedChaotic Data*, Springer, New York, 1996.
- [23] X. He, D. Cai, P. Niyogi. "Laplacian score for feature selection." In *Advances in Neural Information Processing Systems 18* Weiss, Scholkopf, Platt (editors). MIT Press, Cambridge, MA, 2005.