An Anomaly Detection Approach for Dyslexia Diagnosis Using EEG Signals

A. Ortiz[1(✉)], P. J. López[2], J. L. Luque[2], F. J. Martínez-Murcia[1], D. A. Aquino-Britez[3], and J. Ortega[3]

[1] Communications Engineering Department, University of Málaga, Málaga, Spain
aortiz@ic.uma.es

[2] Department of Developmental Psychology, University of Málaga, Málaga, Spain

[3] Department of Computer Architecture, University of Granada, Granada, Spain

Abstract. Developmental dyslexia (DD) is a specific difficulty in the acquisition of reading skills not related to mental age or inadequate schooling. Its prevalence is estimated between 5% and 12% of the population. Currently, biological causes and processes of DD are not well known and it is usually diagnosed by means of specifically designed tests to measure different behavioural variables involved in the reading process. Thus, the diagnosis results depend on the analysis of the test results which is a time-consuming task and prone to error. In this paper we use EEG signals to search for brain activation patterns related to DD that could result useful for differential diagnosis by an objective test. Specifically, we extract spectral features from each electrode. Moreover, the exploration of the activation levels at different brain areas constitutes an step towards the best knowledge of the brain proccesses involved in DD.

Keywords: EEG · Dyslexia · One-Class-SVM · Automatic diagnosis

1 Introduction

Developmental dyslexia (DD) is a specific difficulty in the acquisition of reading skills not related to mental age or inadequate schooling. Its prevalence is estimated between 5% and 12% of the population [7], depending on the reading performance benchmark. It has an important social impact and may determine school failure. In addition, it has harmful effects in the self-esteem of affected children. Early diagnosis and prognosis to start an adequate, early and individualized, intervention is decisive in the in the personal and intellectual development of these children. Currently, biological causes and processes of DD are not well known. It is usually diagnosed by means of specifically designed tests to measure different behavioural variables involved in the reading process. Examples of these variables are reading efficiency, or the ability to split words in their constituent syllables. These tests are individually applied by specialists who need further time to analyze the results and usually, diagnosis is established by means of cut-off points computed over a non very large population. On the other hand,

J. M. Ferrández Vicente et al. (Eds.): IWINAC 2019, LNCS 11486, pp. 369–378, 2019.
https://doi.org/10.1007/978-3-030-19591-5_38

the results of the tests depend on the motivation and the mood of the child when performing the benchmark tasks. As a result, classical diagnosis methods are time-consuming and prone to error, and it is usual that children with specific difficulties in the acquisition of reading skills are neither correctly diagnosed nor treated, what affects their cognitive and emotional development. In addition, most benchmarks are designed for readers, limiting the minimum age for the early diagnosis. Hence, research work oriented towards obtaining results which allow an early diagnosis and an individualized intervention would have a theoretical and a practical impact [10]. There is an active research activity in search of objective, quantifiable measures with diagnostic capability, to improve the diagnosis accuracy and eventually, to reveal unknown aspects of the DD related to its neural basis. Additionally, the research in the biological causes of dyslexia can offer valuable information for a better understanding of the differences between dyslexic and non-dyslexic subjects, with special application in the design of individualized intervention tasks. These quantifiable measures are known as biomarkers, and different studies carried out in the last years used different techniques to extract them. Recent studies searching for DD-related patterns in EEG signals [1,8] have shown differences in readers due to cognitive impairment of the phonological representation of word forms. Speech encoding which is related to speech prosody and sensorimotor synchronization problems can be revealed by finding patterns in EEG channels at different sub-bands as it provides enough time resolution. In this work, we used EEG signals recorded by a 32 active electrodes BrainVision equipment during 5 min sessions, while presenting an auditive stimulus to the subject. These signals are then pre-processed and analyzed in the frequency domain. Spectral features extracted from the EEG signals are then used to classify the subjects between Controls (CN) and Dyslexic (DD).

The rest of the paper is structured as follows. Section 2 shows details on the database used and the applied preprocessing. Then, this section describes the auditive stimulus, EEG preprocessing and post-processing (feature extraction) as well as the classification method. Section 4 presents and discusses the classification results, and finally, Sect. 5 draws the main conclusions.

2 Materials and Methods

2.1 Database

The present experiment was carried out with the understanding and written consent of each child's legal guardian and in the presence thereof. Forty-eight participants took part in the present study, including 32 skilled readers (17 males) and 16 dyslexic readers (7 males) matched in age ($t(1) = -1.4, p > 0.05$, age range: 88–100 months). The mean age of the control group was $94,1 \pm 3.3$ months, and $95,6 \pm 2.9$ months for the dyslexic group. All participants were right-handed Spanish native speakers with no hearing impairments and normal or corrected-to-normal vision. Dyslexic children in this study had all received a formal diagnosis of dyslexia in the school. None of the skilled readers reported reading or spelling difficulties or had received a previous formal diagnosis of dyslexia.

3 Methods

DD is a reading disorder often characterized by reduced awareness of speech units [6]. Recent models of neuronal speech coding suggest that dyslexia originates from the atypical dominant neuronal entrainment in the right hemisphere to the slow-rhythmic prosodic (delta band, 0.5–1 Hz), syllabic (theta band, 4–8 Hz) or the phoneme (gamma band, 12–40 Hz), speech modulations, which are defined by the time of increase in amplitude (i.e., the envelope) generated by the speech rhythm [2,4]. Thus, we compared the cortical entrainment to AM white-noise at a fixed rate in delta (2 Hz), theta (8 Hz) and gamma (20 Hz) bands. In a sample composed of 7 years old children, listened to stimuli obtained by rhythmically modulating the amplitude (AM) of white-noise sound either in the delta, theta and gamma band. Our hypothesis was that the quality of the oscillatory neural processes measured through AM modulations contribute to the optimal construction of predictions of incoming auditory information (such as linguistic sequences or their simplification through AM modulations), these neurophysiological responses should explain the manifestations of the temporal processing deficits described in dyslexia. Then, we recorded EEG signals using a 32 active electrodes (BrainVision actiCAP, https://www.brainproducts.com) while presenting the auditory stimulus. Figure 1 shows the construction of a 8 Hz auditive stimulus, which is based on the AM modulation of bandwidth-limited white noise.

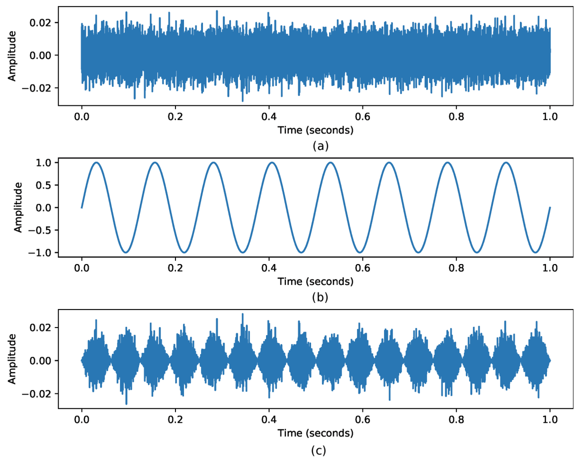


Fig. 1. Stimulus example (a) bandwidth-limited noise, (b) 8 Hz modulating signal, (c) 8 Hz AM Modulated noise

EEG signals were pre-processed in order to remove artefacts related to eye blinking and impedance variations due to movements. Since eye blinking signal is recorded along with EEG signals, these artefact are removed by blind source separation using Independent Component Analysis (ICA) [5]. Other artefacts required the removal of EEG segments. Afterwards, the remaining, cleaned signals were segmented into 5 s excerpts. As a result, a different number of segments are available for different subjects.

Figure 2 shows the average activation levels by frequency bands for the 2 Hz stimulus.

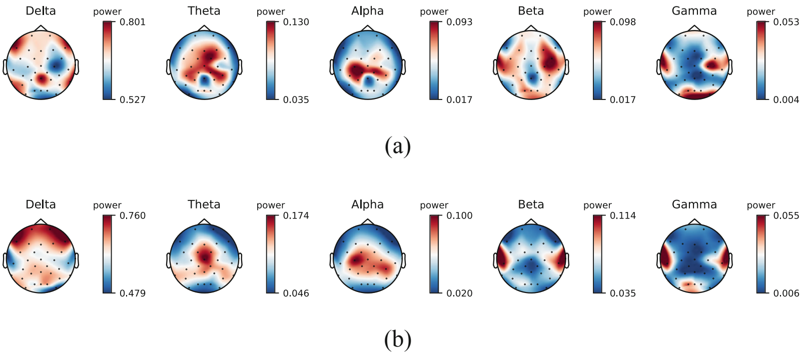


Fig. 2. Average activation patterns computed for 2 Hz stimulus by different bands for (a) Controls and (b) Dyslexic subjects. Multitaper [11] method is used to estimate the PSD.

3.1 Feature Extraction

In this section, we show features extracted from each segment. Since we expect differences in the power spectrum at different frequency bands, we extracted different spectral descriptors. Thus, the first step consist on estimating the Power Spectral Density (PSD). This is usually computed by the Fourier transform. However, the reliability of the PSD computed by this method is reduced by (1) high variance of the estimate, which makes the spectrum noisy and (2) the bias created by the leakage of energy across frequencies [11]. The solution proposed in [11] consist on using windows (also called *tapers*) in the time domain, reducing the leakage produced by multiple side lobes of a window in the frequency domain. This is also achieved by using tapers with low spectral power in the side lobes. Thus, the PSD can be computed as:

$$PSD(\omega) = \left| \sum_{t=0}^{N-1} x(t)a(t)e^{-j\omega t} \right|^2 \tag{1}$$

where $x(t)$ is the N-samples time series corresponding to the signal and $a(t)$ is the window (taper) in the time domain. The total energy of these tappers is normalized to keep the total power invariant. This approach can be extended to reduce the variance of the estimate at each frequency by using multiple tapers. Specifically, Thomson proposed the use of K orthogonal tapers, providing K orthogonal samples of the data $x(t)$. As a result, we have K spectral estimations $PSD_k(\omega)$ that can be averaged to reduce the variance. Furthermore, the method devised by Thomson includes an optimization step to find the tapers that minimize the leakage by maximizing the energy within a specific bandwidth.

Once PSD is computed, two features are extracted to characterize the spectrum of each band for each electrode. The first feature is the *spectral centroid*, (SC) that indicates where the location of the *center of mass* of the spectrum (i.e. the frequency where the PSD is concentrated). This can be calculated as the weighted average of the amplitude spectrum:

$$SC = \frac{\sum_{k=1}^{N} k \cdot w \cdot PSD(k)}{\sum_{k=1}^{N} PSD(K)} \tag{2}$$

where $PSD(k)$ and w are the PSD estimated for the k-bin and the width of each spectral bin, respectively.

Moreover, the mean PSD for each band is also computed and used as a feature.

This way, a feature vector can be composed for the electrode l as

$$f_l = (SC_l^{\Delta}, PSD_l^{\Delta}, SC_l^{\theta}, PSD_l^{\theta}, SC_l^{\alpha}, PSD_l^{\alpha}, SC_l^{\beta}, PSD_l^{\beta}, SC_l^{\gamma}, PSD_l^{\gamma}) \tag{3}$$

for the Delta, Theta, Alpha, Beta and Gamma bands.

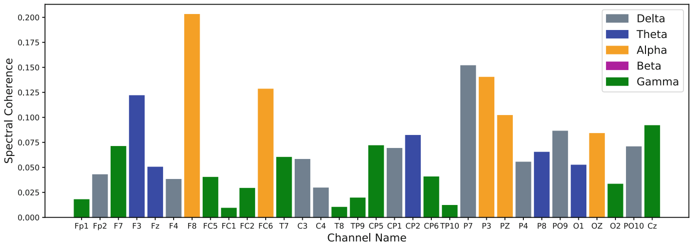
3.2 Feature Selection

Feature selection is addressed by keeping those electrodes presenting a small spectral coherence when comparing Controls and DD. Spectral coherence is a statistic with many applications in neuroscience [3] that measures the relation between the signals acquired from two electrodes $x(t)$ and $y(t)$:

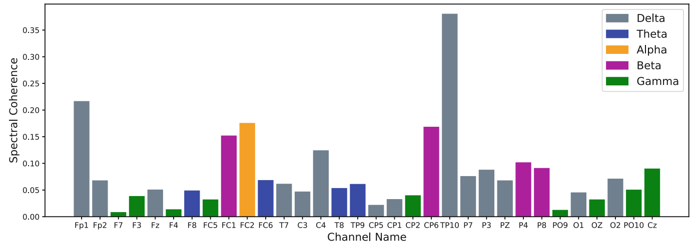
$$C_{xy} = \frac{|C_{xy}|^2}{C_{xx}C_{yy}} \tag{4}$$

where C_{xx} and C_{yy} are the power spectral densities of signals x and y, respectively, and C_{xy} is the cross-spectral density, which can be calculated as the power spectrum of the cross-correlation function between x and y.

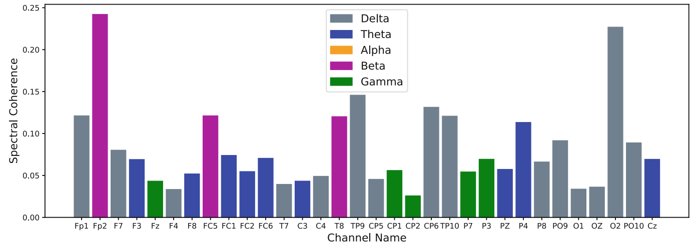
As shown in this figure, different electrodes present different coherence values depending on the frequency band. This indicates that signals acquired by different electrodes contain information regarding different bands. Thus, electrode selection can be addressed by keeping the electrodes that present the lower coherence when comparing CN to DD subjects. Hence, Fig. 3 shows the coherence only for the bands presenting the lowest values.



(a)



(b)



(c)

Fig. 3. Minimum coherence bands for each electrode for (a) 2 Hz stimulus, (b) 8 Hz stimulus and (c) 20 Hz stimulus

3.3 Classification

Class imbalance is an usual problem in biomedical engineering, where databases normally contains more controls than experimental subjects. On the other hand, it is not straightforward to balance the database by obtaining more experimental subjects, due to the distribution of controls and experimental subjects in the general population. As a result, models generated from unbalanced databases are biased, showing special affinity to the most probable class. There are different methods to mitigate the biasing effect such as using cost sensitive objective functions by assigning different weights to miss-classification of samples from different classes. An alternative method to overcome the biasing effect while taking advantage of it consists on modelling the most probable class and then, identify whether a new sample belongs to that distribution or not. This is also known as anomaly detection.

In this work, we used the One-Class SVM [9], a variant of the Support Vector Classifier (SVC) [12] devised to identify outliers with respect to the training dataset. This method separates all datapoints of the training dataset from the origin in the feature space, maximizing the distance from the computed hyperplane to the origin. This is addressed by solving the quadratic programming, minimization problem:

$$\min_{\omega,\xi_i,b} \left\{ \frac{1}{2}\|\omega\| + \frac{a}{\nu N}\sum_{i=1}^{N}\xi_i - b \right\} \tag{5}$$

subject to:

$$\begin{aligned} (\omega \cdot \phi(x_i)) \geq b - \xi_i \qquad & i \in \{1, ..., N\} \\ \xi_i > 0, \nu \in (0, 1] \qquad & i \in \{1, ..., N\} \end{aligned}$$

where ξ_i are non-zero variables to control the margin, and ν controls the number of support vectors and the fraction of training samples considered as outliers. Additionally, ϕ is the kernel function.

Hence, a decision function can be constructed to produce a different value for samples belonging to the same distribution of the training samples that for out-of-class samples, using the hyperplane defined by ω and b parameters

$$f(z) = sign\{(\omega \cdot \phi(z)) - b\} \tag{6}$$

In our experiments, a Radial Basis Function was used for the kernel.

4 Results and Discussion

In this section, we present the experimental results obtained when classifying the subjects by means of the features extracted from the EEG signals. These classification experiments used EEG features from signals acquired during the 2 Hz, 8 Hz and 20 Hz stimuli as explained in Sect. 3. Moreover, experiments using all the features and the selection provided by the method described in Sect. 3.2 are shown here. The classification method exposed here has been assessed by stratified k-fold cross-validation (k = 5) to ensure the database independence and to avoid double dipping in the training-testing process.

Thus, Figs. 4a, b and c, shows the ROC curves obtained when classifying the subjects using th 2 Hz, 8 Hz and 20 Hz stimuli, respectively.

The feature selection method based on using only the band that shows the lowest coherence between CN and DD subjects, improves the performance of the classifier with respect to the use of all the features for 2 Hz and 8 Hz. The improvement of the performance comes from the reduction of the dimensionality and the use of more discriminative features. Nevertheless, the use of all

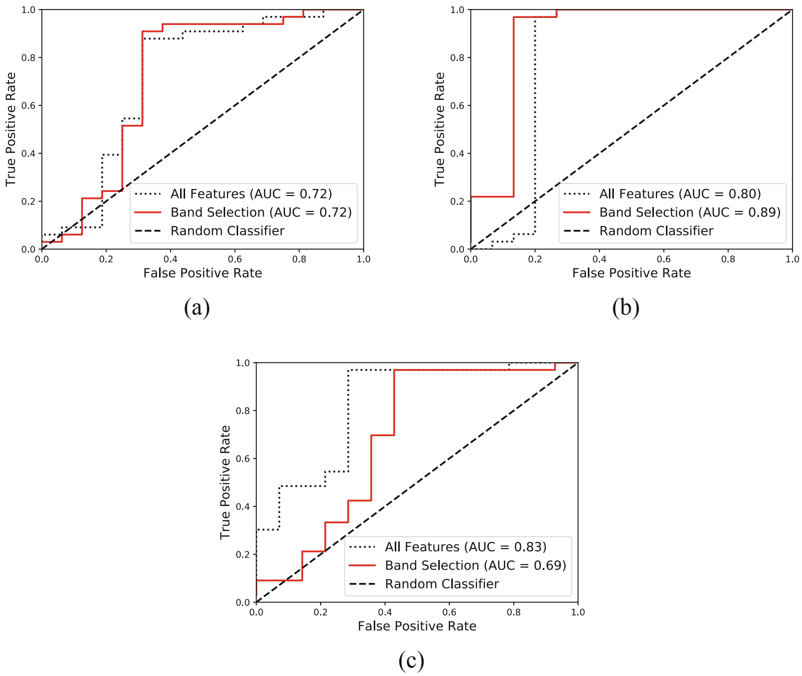


Fig. 4. ROC curves obtained with the (a) 2 Hz, (b) 8 Hz and (c) 20 Hz stimuli. All features vs. per electrode band selection method is shown.

the features (i.e. all bands for all the electrodes) provides higher AUC values for the 20 Hz stimulus. Moreover, Table 1 shows the classification performance in terms of accuracy, sensitivity and specificity. As shown in this table, the feature selection method improves the sensitivity and specificity for 2 Hz and 8 Hz, while decreases the performance in the 20 Hz case. This suggest that discriminative information regarding electrode inter-dependence is present when the 20 Hz stimulus is used.

Table 1. Classification results

Stimulus	Accuracy	Sensitivity	Specificity	AUC
2 Hz (All features)	0.62	0.66	0.60	0.72
2 Hz (Band Selection)	0.70	0.66	0.69	0.72
8 Hz (All Features)	0.63	0.80	0.55	0.80
8 Hz (Band Selection)	0.66	0.86	0.56	0.89
20 Hz (All Features)	0.78	0.66	0.81	0.83
20 Hz (Band Selection)	0.71	0.53	0.78	0.69

5 Conclusions and Future Work

In this paper, EEG signals have been recorded during the presentation of different stimulus related to the frequency of neural oscillations generated at different brain areas during language processing. Then, a feature extraction process directed to characterize the signals from each electrode in terms of the predominant brainwave. Moreover, these features are selected by computing the spectral coherence for all electrodes between controls and experimental subjects. The feature extraction and selection method used in this work improves the classification performance for 2 Hz and 8 Hz stimulus, which suggest that discriminative information regarding DD diagnosis is in the distribution of power along different bands. In fact, the proposed method always provides AUC values up to 0.89, showing its diagnostic utility. In addition, the 20 Hz seems to produce effects beyond the spectral distribution and thus, a different feature selection method has to be used. In a future work, we will explore the use of different, time-frequency features and different descriptors to characterize the power distribution along different bands, as well as to compute electrode synchronicity among different brain areas.

Acknowledgments. This work was partly supported by the MINECO/FEDER under PSI2015-65848-R and TEC2015-64718-R projects.

References

1. Cutini, S., Szũcs, D., Mead, N., Huss, M., Goswami, U.: Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia. NeuroImage **143**, 40–49 (2016)
2. Di Liberto, G., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., Lalor, E.: Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. NeuroImage **175**, 70–79 (2018)
3. Engel, A.K., Fries, P., Singer, W.: Dynamic predictions: oscillations and synchrony in top-down processing. Nat. Rev. Neurosci. **2**, 704–716 (2001)
4. Flanagan, S., Goswami, U.: The role of phase synchronisation between low frequency amplitude modulations in child phonology and morphology speech tasks. J. Acoust. Soc. Am. **143**, 1366–1375 (2018). https://doi.org/10.1121/1.5026239
5. Li, R., Principe, J.C.: Blinking artifact removal in cognitive EEG data using ICA. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5273–5276 (2006)
6. Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M.: Out-of-synchrony speech entrainment in developmental dyslexia. Hum. Brain Mapp. **37**, 2767–2783 (2016)
7. Peterson, R., Pennington, B.: Developmental dyslexia. Lancet **379**, 1997–2007 (2012)
8. Power, A.J., Colling, L., Mead, N., Barnes, L., Goswami, U.: Neural encoding of the speech envelope by children with developmental dyslexia. Brain Lang. **160**, 1–10 (2016)

9. Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. In: Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS 1999, pp. 582–588. MIT Press, Cambridge (1999)
10. Thompson, P.A., Hulme, C., Nash, H.M., Gooch, D., Hayiou-Thomas, E., Snowling, M.J.: Developmental dyslexia: predicting individual risk. J. Child Psychol. Psychiatry **56**(9), 976–987 (2015)
11. Thomson, D.: Spectrum estimation and harmonic analysis. Proc. IEEE **70**, 1055–1096 (1982)
12. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)