An Auditory Brainstem Response-Based Expert System for ADHD Diagnosis Using Recurrence Qualification Analysis and Wavelet Support Vector Machine

Zeynab Esmailpoor
Biomedical Engineering Department,
Amirkabir University of Technology,
Tehran, Iran
esmailpoor@aut.ac.ir

Ali Motie Nasrabadi
Engineering Department,
Shahed University,
Tehran, Iran
nasrabadi@shahed.ac.ir

Saeed Malayeri
University of Social Welfare and
Rehabilitation Science,
Tehran, Iran
malayeris@gmail.com

***Abstract*:**

Attention Deficit Hyperactivity Disorder (ADHD) is a common disorder in children. Due to lack of suitable biomarker or test, diagnosis of ADHD children is complicated and needs comprehensive evaluations. Evidences show that, ADHD children have deficit in their brainstem timing and cortex auditory processing. We assessed their auditory brainstem response to speech stimuli. Due to nonlinear and dynamic characteristics of biological signals they should be represented by features that are based on their nature. In this study wavelet coefficients and recurrence qualification analysis features were used to represent signals in a comprehensive way. In this article, we addressed the problem of discrimination of ADHD children from Normal. Wavelet Support Vector machine with Mexican hat and Morlet kernels were used in order to classifying these children. Our method demonstrated %98.57 classification accuracy.

Keywords: Support Vector Machine (SVM), ADHD children, Recurrence Qualification Analysis (RQA).

I. INTRODUCTION

The Auditory Brainstem Response (ABR) to speech provides a way to understand the subcortical auditory processing mechanisms. Deficit in sound encoding that is associated with learning and auditory processing disorder may be diagnosed by ABR as a biological marker [1]. ABR can be considered as the early portion of the Auditory Evoked Potentials (AEPs) and reflect the processing of acoustic stimuli in the brainstem [2]. The ABR to speech copies the acoustic characteristics of the speech stimuli with remarkable precision. This valuable property makes it possible to derive clinically applicable information related to the auditory processing of its complex stimuli [1].

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common psychiatric disorders of childhood [3]. Although ADHD has been viewed as a neurological disorder, the relation between the ADHD features and its counterpart underlying neurocognitive disorder is not that clear [4]. Due to involving a wide range of attention associated neural networks in cognitive processes, the identification of deficits of ADHD patients is really complicated [5]. Likewise, due to lack of suitable test or biomarker for diagnosis of ADHD children their identification requires a complete evaluation of children's behavior both at home and at school. As a result, their diagnosis would be somewhat less accurate and time-consuming. By taking the impacts of ADHD on the future lives of its suffering children into account, there is an absolute need to a biological marker which is capable enough to diagnose the ADHD children at the beginning of observing its symptoms in childhood. In order to introduce working biomarkers, some valuable research works have been conducted about Electroencephalography (EEG) signals and ABR for diagnosis of ADHD children during the past decades [4, 6].

Evidences show that, automatic recognition and classification of the biomedical signals can present working results since they have low computational complexity and high accuracy. Consequently, the automatic analysis of the bioelectric signals can considerably reduce the mistakes which usually occur in the analysis of the physicians and medical experts. Analyzing the EEG signal by using Wavelet signal processing technique and neural network was used to classify ADHD from Normal in one study [7]. They obtained 95.6% classification accuracy by using radial basis function neural network as a classifier.

Support vector machine for classifying the ADHD patients was used by Muller et. al [8]. They benefited from Independent Component Analysis (ICA) where the ERPs were decomposed into functionally different components. Results show that, the classification accuracy of their method regarding to using 10-fold cross-validation strategy is about 92 percent. In another study the EEG measurement in four different conditions such as open and close eyes and two neuropsychological tasks (visual and emotional continuous performance tests) was utilized in order to discriminate between ADHD patients and normal groups. By utilizing support vector machine (SVM) they obtained maximum 82.3% classification accuracy [9].

In spite of using the EEG signals for classifying the ADHD children, there is no published report about using the ABR signals.

We suggested to use Speech-ABR as a biomarker because not only recording Speech-ABR compared with EEG signal is much easier, but also it can concurrently assess the hearing ability. The goal of this study is proposing an efficient system for detection and classification of ADHD patients. The organization of the paper will be as follows. The first section consist of the experiment participants, data acquisition, a methods for feature extraction and classification. Results and discussion are being presented in the third section and finally conclusion is given in the fifth part.

II. PROPOSED METHOD

A. Participants

The 67 Persian speaking children between the ages of eight and 12 years participated in our experiment. All the children had normal hearing ability. The ADHD patients were diagnosed in

978-1-4799-1972-7/15/$31.00

a clinical interview with their parents by children psychiatrist according to DSM-IV criterion [10]. The control group was selected from normal school children having no symptoms of ADHD according to DSM-IV criteria. Control group consists of 35 children (20 boys and 15 girls) and also the ADHD group consist of 31 children (14 girls and 17 boys).

B. Data Aquisition

ABR were elicited by an acoustic click and a speech syllable /da/, and both brainstem responses were collected in the same manner and during the same recording session using the biomarker Navigator Pro System (Natus Medical Inc, Mundelein,USA). Responses were recorded from Ag-AgCl electrodes, with contact impedance of <5kΩ , positioned on the forehead (active), behind the right ear (ground) and behind left ear lobe (reference). Stimuli were presented into the right ear at 80.3dbSPL through insert earphones. The Sampling rate was 12 kHz. These response are elicited passively, subjects are not attending to stimuli.
The speech ABR was elicited using a five-formant speech syllable 40ms /da/ stimulus with an onset burst during the first 10ms (figure1). One block of 4000 sweeps with the rate of 10.9/sec were presented and the speech evoked responses were averaged online. The recording window was 85.33 ms starting 15ms prior to stimulus onset and 20ms posterior to stimulus.

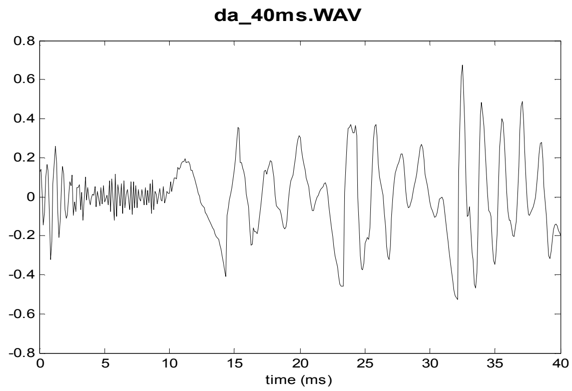

Figure 1: Stimulus waveform of 40ms speech /da/

The brainstem response to /da/ speech syllable is made up of two separate elements, the onset response and the Frequency-Following Response (FFR). For a syllable /da/, the onset response corresponds with the burst release of stop consonant (/d/), the FFR response reflects the transition period between the burst and the onset of the vowel itself. A response of a normal child to a 40 ms syllable /da/ is shown in Figure2.

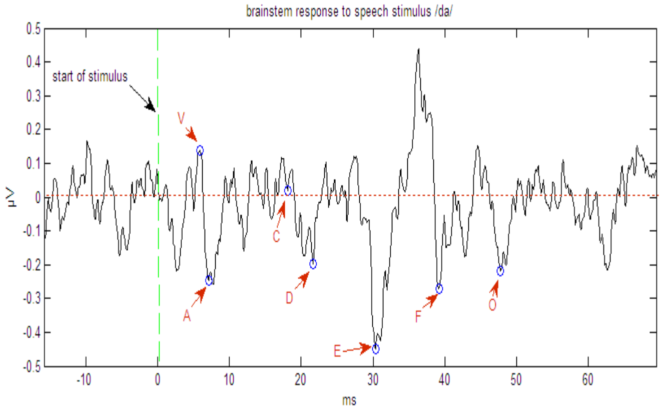

Figure 2: Response to speech stimulus /da/.

C. Feature extraction

Art of signal processing is to extract features that are able to describe signal completely in a way that we have minimum lost information. In this study, firstly we decided to use wavelet coefficients because of non-stationary nature of biological signals.

1) Wavelet analysis for feature extraction

Wavelet is a powerful technique for analyzing and multi-scale representation of the given signal. The Discreet Wavelet Transform (DWT) is the most widely used algorithm for multi-resolution AEP analysis[11]. Varying window size in wavelet transform (WT) helps to have optimum resolution in high and low frequencies. In this study, biorthogonal 5.5 (Bior 5.5) was chosen as the mother function of WT and it was computed for each subject in ADHD and Normal group separately. Signals were evaluated up to fifth scale. Five detail scales ($D_1 - D_5$) and final approximation (A_5) with different numbers of wavelet coefficients at each scale are obtained. In each scale coefficients which can cause meaningful discrimination are chosen ($p<0.05$). So, 19 coefficients of D5, 24 coefficients of D4, 39 coefficients of D3, 69 coefficients of D2 and 64 coefficients of D1 are selected. Totally we have 215 wavelet coefficients that can discriminate ADHD from Normal children significantly. Beside the time-frequency features of signal that may represent the signal, it would be better to consider non-linear and dynamical feature of signal too, due to dynamic and non-linear behavior of biological signals.

2) Recurrence Qualification Analysis

It has been believed that in deterministic dynamical systems and also in nonlinear and chaotic ones, states of system will be approximately repeated after a while. It is a fundamental property of these systems [12, 13]. The method of recurrence plots (RP) was introduced to visualize the time dependent behavior of the dynamic of systems, which can be pictured as a trajectory in the phase space[13]. RP can be mathematically expressed as:

$$R_{i,j}^{m,\varepsilon_i} = \Theta\left(\varepsilon_i - \left\|\vec{x_i} - \vec{x_j}\right\|\right), \ \vec{x_i} \in R^m , \qquad i,j = 1 \ldots . N$$

Where N is the number of considered states $\vec{x_i}$, ε_i is a threshold distance, $\|.\|$ a norm and $\Theta(.)$ the Heaviside function. Usually the phase space has to be reconstructed from the original one directional time series [14]. Because analysis of systems trajectory in phase space is dependent on the embedding dimension, m, it has to be chosen appropriately.

In recurrence plots there are three small structures: single points which can occur if states are rare; a diagonal line of length $l(R_{i+k,j+k} = 1 \Big|{l-1 \atop k=0})$, occurs when a segment of the trajectory runs almost in parallel to another segment, and a vertical (horizontal) line with v the length of the vertical line($R_{i,j+k} = 1 \Big|{v-1 \atop k=0}$), markes a time interval in which a state does not change or change very slowly.

In order to go beyond the visual impression yielded by RPs, several measures of complexity which quantify the small structures in RPs have been proposed in some studies [15] and are known as recurrence qualification analysis (RQA). A computation of these measures in small windows (sub-matrices) of the RP moving along the LOI[1] yields the time dependent behavior of these variables. Recurrence rate (RR), based on the recurrence point density, is simply the average number of neighbors that each point in the trajectory has in its ε-neighborhood. The measures based on diagonal line distribution are *determinism* (DET), Average diagonal line length ($L_{mean}\ or < L >$), the longest diagonal line (L_{max}) and entropy (ENTR). The diagonal line distribution encodes main properties of the system, such as predictability and measures of complexity. The more a system is determined, the greater amplitude of these measures. Finally measures based on vertical lines structures are *laminarity* (LAM), *trapping time* (TT), and *maximal length of the vertical lines* (v_{max}). Measures based on vertical lines, marks state where trapped for some time. This is a typical behavior of laminar states [15-17].

In this study we considered $v_{min} = l_{min} = 2$, moving window size $w = 8ms$, and window step size 0.08ms. Embedding dimension was calculated by *false nearest neighbor* (FNN) method and delay was selected by *mutual information* (MI) approach. RQA computation was performed for both ADHD and Normal children's ABR signals. So, we have nine recurrence features for each child. For each of these nine features by using moving window along the LOI, time dependent behavior of these features was achieved. So, Minimum, maximum, variance and mean of each feature was computed. Totally we extracted 36 nonlinear features that only 20 of them could statistically discriminate ADHD from Normal children.

D. Classification

As we know, not only a strong classifier should have a deep connection with the data distribution, but also could be easily generalizable. Due to nonlinear distribution of the biological features, choosing appropriate classifier is critical. The W-SVM is relatively new and powerful technique for solving supervised classification problem and is very useful due to its generalization ability. In essence, such an approach maximizes the margin between the training data and the decision boundary, which can be cast as a quadratic optimization problem. The subsets of patterns that are closest to the decision boundary are called support vectors[2].

SVM uses a device called kernel mapping to map the data in input space to a high-dimensional feature space in which the problem becomes linearly separable[18]. The decision function of an SVM is related not only to the number of SVs and their weights but also to the a priory chosen kernel that is called the support vector kernel[18, 19]. Many kinds of kernels can be used, such as Gaussian and polynomial kernels. In theory, wavelet decomposition emerges as a powerful tool for approximation that is to say the wavelet function is a set of bases that can approximate arbitrary functions. Wavelet kernel has the same expression as a multidimensional wavelet function; therefore the goal of WSVM is to find the optimal approximation or classification in the space spanned by multidimensional wavelets or wavelet kernels[20].

In this study, wavelet support vector machine (WSVM) classifier is used. Morlet and Mexican hat wavelet kernels are used in this study. The wavelet kernels are shown in table1 [21].

Table 1: Wavelet kernel for WSVM

	Wavelet kernel function
Morlet	$k(x,x') = \prod_{i=1}^{d} \cos\left(w_0 \frac{(x_i - x_i')}{a_i}\right).\exp(-\frac{\|x_i - x_i'\|^2}{2a_i^2})$
Mexican hat	$k(x,x') = \prod_{i=1}^{d} (1 - \left(\frac{(x_i - x_i')^2}{a_i^2}\right)).\exp(-\frac{\|x_i - x_i'\|^2}{2a_i^2})$

Parameters related to each kernel have been chosen based on trial and error.

III. RESULTS AND DISCUSSION

In this study, auditory brainstem responses were evaluated form two different points of view; each signal was described by both time-frequency components and nonlinear behaviors in phase space. As mentioned, we obtained 215 wavelet coefficients and 20 RQA features that could discriminate ADHD and normal group significantly. Thus we gained 235 features. In order to extract an optimum subset of features, stepwised method of SPSS software was used. So, in wavelet features 37 of 215 was selected as an optimum subgroup and totally 27 features of 235 features.

In classification approach, dataset should be divided into training and test subgroups. As we had 31 ADHD and 37 Normal children we used K-fold (5-folds) Cross Validation in which all data were randomly divided into five groups. Each time 4 groups were used for training and the last one for test.

In classification phase, wavelet features, RQA features and combined features are classified separately. The results of classification are shown in table 2,3.

[1] Line of Identity

Table 2: results of SVM classification by Mexican hat kernel

features	Mexican hat kernel		
	Sensitivity	specificity	Accuracy
RQA features ($\sigma = 1$)	87%	83%	85.05±0.13
Wavelet features ($\sigma = 8$)	97%	97%	97.13±0.04
Combination of wavelet and RQA ($\sigma = 5$)	100%	97%	98.57±0.03

Table 3: results of SVM classification by Morlet kernel

features	Morlet kernel		
	Sensitivity	specificity	Accuracy
RQA features $(w_0, \sigma) = (5, 2)$	83%	91%	88.13±0.07
Wavelet features $(w_0, \sigma) = (2, 10)$	97%	94%	95.6±0.06
Combination of wavelet and RQA $(w_0, \sigma) = (3, 2)$	100%	97%	98.58±0.03

Classification results demonstrated that combination of these two kinds of features represent the signal in a better way. Results show that there is approximately no difference between accuracy of Morlet and Mexican hat kernel classification but combination of all features significantly affects the accuracy of classification.
To the best of our knowledge, this is the first work trying to classify ADHD by using ABR signals. Retrospectively, EEG has been used to discriminate ADHD children from normal [6, 8]. Their results revealed acceptable accuracy regarding difficulty and complexity of recording EEG signal. On the other hand, ABR could be recorded easier and simultaneously asses hearing ability of children.

IV. CONCLUSION

ADHD is one of the most common disorders in children that its impacts on the future of its suffering is remarkable. Evidences show that, ADHD children have deficit in their brainstem timing and cortex auditory processing. In this study we hypothesized that ABR signals may assist us in diagnosing ADHD children. Biological signals such as auditory brainstem responses are naturally dynamic [22], therefore time and frequency domain features cannot describe it completely. In this study signals were analyzed from two different perspectives. Features extracted from phase space and also time-frequency features were combined in order to represent the signals in a comprehensive way. By considering the results we can conclude that ABR to speech could be a biomarker for diagnosis of ADHD in children. More to the point, Automatic analysis of signals considerably reduce mistake that may occur in evaluating child's behavior and also has a high accuracy.

V. REFERENCES

[1] l.Johnson, K., T. G.Nicol, and N. Kraus, *Brain Stem Response to Speech:A Biological Marker of Auditory Processing.* Ear & Hearing, 2005. **26**: p. 424-434.

[2] Acir, N., O. Ozdamar, and C. Guzelis, *Automatic classification of auditory brainstem responce using SVM-based feature selection algorithm for threshold detection.* engineering Applications of Artificial Inteligence, 2006. **19**: p. 209-218.

[3] G. Polanczyk, et al., *The worldwide prevalence of ADHD:A systematic review and metaregression analysis.* The American Journal of Psychiatry, 2007: p. 942-948.

[4] Azzam, H. and D. Hassan, *Auditory Brainstem Timing and Cortical Processing in Attention Deficit Hyperactivity Disorder.* Current Psychiatry, 2010.

[5] solomon, M., S.J. Ozonoff, and S.Ursu, *The neural substraites of cognitive control deficits in autismspectrum disorders.* Neuropsychologia, 2009: p. 2515-2526.

[6] M.Ahmadlou and H.Adeli, *wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD.* 2010.

[7] Ahmadlou, M. and H. Adeli, *Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD.* Clinical EEG and Neuroscience, 2010. **41**(1): p. 1-10.

[8] Mueller, A., et al., *Classification of ADHD patents on the basis of independent ERP components using a machine learning system.* Nonlinear Biomedical Physics, 2010. **4(supple 1)**.

[9] Tenev, A., et al., *Machine learning approach for classification of ADHD adults.* International Journal of Psychophysiology, 2013.

[10] Association, A.p., *diagnostic and statistical manual of mental disorders, forth edition, text revision*. 2000, washington D.C.

[11] Zhang, R., et al., *Classification of the Auditory Brainstem Response(ABR) using Wavelet Analysis and Baysian Network.* Procceding of the 18th IEEE symposium on Computer-Based Medical System(CBMs05), 2005.

[12] Ott, E., *choas in dynamical systems*. 1995, university press,cambridge.

[13] Eckmann, J.-P., S.-O. Kamphorst, and R. D, *Recurrence plots of dynamic systems.* Europhys, 1987: p. 973-977.

[14] Takens, F., *Detecting strange attractors in turbulence.* springer,Lecture notes in mathematics, 1981: p. 366-381.

[15] Marvan, N., M. A, and J.Bifur, *extended recurrence plot analysis and its application to ERP data.* Chaos Cogn Int Complex Brain Dynamic, 2002: p. 761-771.

[16] Talebi, N. and A.M. Nasrabadi, *Investigation of changes in EEG complexity during memory retrivial:the effect of midazolam.* Cogn Neurodyn, 2012: p. 537-546.

[17] Theiler, *spurious dimention from correlation algorithms applies to limited time series data.* Phys Rev, 1986. **34**: p. 2427-2432.

[18] Burges, C.J.C., *A tutorial on support vector machines for pattern recognition.* Data mining and knowledge Discovery, 1998. **2**: p. 121-167.

[19] A.Smola, B. Scholkopf, and K.R. Muller, *The connection between regularization operators and support vector kernels.* Neural network, 1998d. **11**: p. 637-649.

[20] Zhang, L., W. Zhou, and L. jiao, *Wavelet Support Vector Machines.* IEE Transaction on Systems, Man and Cybernetics, 2004. **34**.

[21] cui, w.z., *wavelet support vector machine with universal pproximation and its application*, in *IEEE information theory workshop*. 2006. p. 340-364.

[22] Chandrasekaran, B. and N. Kraus, *The scalp-recorded brainstem response to speech: Neural origins and plasticity.* Psychophysiology, 2010. **47**(2): p. 236-246.