Clinical Neurophysiology 127 (2016) 2695–2703

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Automatic classification of 6-month-old infants at familial risk for language-based learning disorder using a support vector machine

Marzieh Zare [a,b], Zahra Rezvani [b], April A. Benasich [c,*]

[a] School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran 19538, Iran
[b] Institute for Cognitive and Brain Sciences, Shahid Beheshti University, G.C. Evin, Tehran 19393, Iran
[c] Center for Molecular and Behavioral Neuroscience, Rutgers University-Newark, 197 University Avenue, NJ 07102 Newark, USA

See Editorial, pages 2692–2694

ARTICLE INFO

Article history:
Accepted 25 March 2016
Available online 12 April 2016

Keywords:
Infant
Machine learning
Support vector machine (SVM)
EEG
Developmental language disorder
Network analysis

HIGHLIGHTS

- Novel machine learning approaches were used to study selected features within infant resting EEG.
- Two infant groups who differed on familial risk for language learning disorder (LLD) were assessed.
- Identification of infants at higher risk for LLD may facilitate earlier diagnosis and remediation.

ABSTRACT

Objectives: This study assesses the ability of a novel, "automatic classification" approach to facilitate identification of infants at highest familial risk for language-learning disorders (LLD) and to provide converging assessments to enable earlier detection of developmental disorders that disrupt language acquisition.
Methods: Network connectivity measures derived from 62-channel electroencephalogram (EEG) recording were used to identify selected features within two infant groups who differed on LLD risk: infants with a family history of LLD (FH+) and typically-developing infants without such a history (FH−). A support vector machine was deployed; global efficiency and global and local clustering coefficients were computed. A novel minimum spanning tree (MST) approach was also applied. Cross-validation was employed to assess the resultant classification.
Results: Infants were classified with about 80% accuracy into FH+ and FH− groups with 89% specificity and precision of 92%. Clustering patterns differed by risk group and MST network analysis suggests that FH+ infants' EEG complexity patterns were significantly different from FH− infants.
Conclusions: The automatic classification techniques used here were shown to be both robust and reliable and should provide valuable information when applied to early identification of risk or clinical groups.
Significance: The ability to identify infants at highest risk for LLD using "automatic classification" strategies is a novel convergent approach that may facilitate earlier diagnosis and remediation.

1. Introduction

The brain is subject to large structural and functional changes over early development. Rapid auditory processing and auditory change detection abilities in the tens-of milliseconds range are critical to decoding the speech stream and are crucial aspects of speech and language development starting at birth (Aslin, 1989; Eilers et al., 1981; Werker and Tees, 2005; Kuhl et al., 2008). This complex ability to detect subtle sound changes early in infancy, specifically

* Corresponding author at: Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Avenue, Newark, NJ 07102, USA. Tel.: +1 973 353 3598; fax: +1 973 353 1760.
E-mail addresses: marziehzare@ipm.ir (M. Zare), za.rezvani@mail.sbu.ac.ir (Z. Rezvani), benasich@andromeda.rutgers.edu (A.A. Benasich).

http://dx.doi.org/10.1016/j.clinph.2016.03.025
1388-2457/

fast sequential changes in the amplitude and frequency composition related to speech, is believed to go awry in a subset of children (Benasich and Tallal, 2002; Choudhury and Benasich, 2011; Tallal, 2004). Such early deviation from normative acoustic processing trajectories is thought to result in language-based learning disorders (LLD) such as specific language impairment and dyslexia (Bush, 2010; Lewis and Elman, 2008; Tallal, 2004), and is suggested to be comorbid with some types of autism (Whitehouse et al., 2008). Familial genetic studies indicate that approximately 30–60% of infants born into families with LLDs are at risk of developing similar problems (Flax et al., 2003; Tomblin, 1989).

Previous studies also indicate that LLDs are associated with detectable differences in brain structure (Chu et al., 2015; Leonard et al., 2011; Westwood, 2004) that may begin before birth. These anatomical differences are thought to be a contributing factor to LLD given the genetic predisposition identified in many of these disorders (Casey et al., 2000; Choudhury and Benasich, 2011; Wong et al., 2013). Therefore, prospective longitudinal studies starting in early infancy and continuing through 3–5 years of age have been designed to detect early precursors and biomarkers of developmental disorders in infants at higher genetic risk of LLD (for review see Benasich and Choudhury, 2012). These early risk markers are difficult to detect using only behavioral testing, however neuroimaging approaches, including EEG, have been effectively used in infant populations (e.g. Benasich et al., 2006; Choudhury and Benasich, 2011; Maitre et al., 2013).

The recent focus has been on improving event-related potential (ERP) recording from EEG and magnetoencephalography (MEG) as well as introducing more fine-grained analyses of continuous EEG, particularly within the context of studying atypical or at-risk groups (Barttfeld et al., 2011; Bosl et al., 2011; Stahl et al., 2012). However, the standard procedure of EEG and MEG analysis continues to be averaging of a large number of artifact-free trials and then using group grand averages to compute statistics. Unfortunately, this can result in a number of problems given the underlying assumptions of this technique. Violation of these basic assumptions arises due to inconsistency of the brain response, variability across trials due to cognitive processes and loss of statistical power as well as statistical bias due to major alterations in the ERP components (Stets et al., 2012). Moreover, in developmental studies, a large proportion of subjects may need to be excluded because the infants have not provided a sufficient number of noise-free trials per condition, thus precluding computation of stable averaged ERPs.

Another approach has focused on using source localization and time/frequency analyses in resting or spontaneous EEG to identify predictors in at-risk populations (e.g. Benasich et al., 2008; Gou et al., 2011) as well as those already diagnosed with a LLD (e.g. Heim and Benasich, 2011; Schiavone et al., 2014). These studies examine particular frequency profiles using Fast Fourier Transform (FFT) or wavelet analyses, thus using more complex oscillatory characteristics. This technique overcomes some of the difficulties of averaged ERPs, however, the information obtained is restricted to the oscillatory domain and must be first computed on a case-by-case basis and then averaged to obtain group data.

In order to improve statistical power and avoid the issues that emerge in MEG, EEG and ERP studies it would be advantageous to have novel statistical methods that would permit *detailed discrimination of individual characteristics* (using raw or minimally pre-processed EEG data) as well as supporting robust classification of groups that may differ on the level of risk for a particular disorder (Stahl et al., 2012).

Automatic classification strategies using machine-learning classifiers have been suggested as just this type of diagnostic tool (Riaz et al., 2013). Machine learning techniques enable leverage of widely distributed, but potentially less robust information to improve the ability to separate individuals into risk groups. Analytic approaches such as FTT do not support group classification and are restricted to features within the oscillatory domain. However, the automatic classification strategies described here use many different features, including oscillatory characteristics, which are driven by the networks detected in the EEG or MEG signal. For example, in one study a multivariate pattern classification approach was applied to network based fcMRI data from a large and unique infant study sample processed with current motion correction procedures (Pruett et al., 2015); significant changes were demonstrated in the structure of large-scale functional brain networks over 6- to 12- months, a period of dramatic cognitive, motor, and social transformation. Hence, this early time period was shown to hold great importance for understanding typical and atypical social-developmental trajectories (Elison et al., 2013). In the present dense-array EEG study, we focus on network connectivity of two infant groups with differing LLD risk levels and deploy a machine learning classifier called a support vector machine (SVM) to separate and classify these infants at low and high risk of LLD. Subjects were two groups of infants: *Family History Positive* (FH+) *group*: Infants born into families with a history of LLD and *Family History Negative* (FH−) *group*: Typically developing control infants without such a family history. We set out to address (i) whether the brain network topology is affected by degree of LLD risk, (ii) which features derived from network analysis contribute to classifications (iii) whether a classifier for prediction of impairment can be implemented to aid early risk assessment of LLD.

2. Participants

Participants in the current study were a subset of the children who participated in a larger prospective study that assessed the effects of early auditory processing skills on later language and cognitive development (Benasich et al., 2006; Choudhury and Benasich, 2011; Choudhury et al., 2007). The studies described here have been reviewed and approved by the Institutional Review Board of our University, were in accordance with the ethical standards of our University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all parents following a full explanation of the experiment and prior to their child's inclusion in the reported studies. All children were tested at 6, 9, 12, 16, 24, 36 and 48 months of age using both behavioral information processing tasks and EEG assessments. Results from the behavioral information processing assessments are presented elsewhere (Choudhury et al., 2007) as are the analysis of the EEG event-related responses (ERPs) to rapidly modulated auditory stimuli (Choudhury and Benasich, 2011). In the present study we focus only on the 6-month-old infants in the two risk groups: FH+ and FH−).

Families were recruited from urban and suburban communities in New Jersey and assigned to one of the two groups based on parental report of family history of LLD. The FH+ group consisted of 12 full-term normal birth weight healthy infants (10 males, 2 females). Infants from FH− families were recruited from local newspaper birth announcements and pediatric clinics (7 males, 5 females). In order to be classified as FH+, families were asked to provide clinical reports of expressive and receptive language scores and a general cognitive score for at least one affected and diagnosed immediate family member (the "proband"); 75% of the probands were siblings of the infant participant and the remainder were parents who had been clinically diagnosed with a LLD. All probands for this sample had diagnoses of either specific language impairment or developmental dyslexia. (Further information

regarding ascertainment may be found in Choudhury and Benasich, 2011 and Choudhury et al., 2007. Discussion of the distribution of genders in families with an LLD proband can be found in Choudhury and Benasich, 2003.)

The electroencephalogram (EEG) was recorded from the subjects while they were awake, alert and seated comfortably on their parent's laps in a sound-attenuated and electrically shielded chamber. Silent videos were played on a monitor in front of the children to engage them and minimize their movement. An experimenter engaged the children attention with a silent puppet show or other toys if they lost interest in the video. EEG was recorded from 62 scalp sites using the Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR) in order to examine changes in each child's spontaneous brain electrical activity. The vertex was used as the online reference electrode. The signal was sampled at 250 Hz and bandpass filtered online at 0.1–100 Hz.

3. Methodology

Total duration of the continuous resting state data collected from each child was approximately three minutes. Channels containing noise and artifacts were visually inspected and removed by an investigator. The continuous EEG data was split into epochs (mean = 27,590 samples or ~2 min) for each child. Although epochs were slightly variable in length, the variability was fairly small (27,500–35,000) and did not differ between groups. Using automatic channel rejection in EEGLAB (Delorme and Makeig, 2004), noisy channels were identified, removed and interpolated. Channels 63 and 64 are placeholders in EGI geodesic infant nets and are automatically removed from analyses. The artifact-free epochs were submitted to power spectral analysis using Fast Fourier Transform (FFT). The log transform for absolute power was: 10 log10(x). Time series are broken into six frequency bands [delta (0.5–4 Hz); theta (4–8 Hz); alpha1 (8–10 Hz); alpha2 (10–13 Hz); beta (13–30 Hz); and gamma (30–48 Hz)].

3.1. Graph Analysis

For graph analysis, connectivity matrices were constructed based on correlation analyses. (See Fig. 1 for an overview of the graph analysis methods used here.) The connectivity matrix defines a weighted graph where each electrode corresponds to a node and the weight of each link is determined by correlation of the electrode pair. In this study the weights of connection are derived from the correlation matrix where the elements are ρ_{ij}, i and j are different channels of our EEG data. The correlation matrix is calculated using the following equation (Rodgers and Nicewander, 1988):

$$\rho_{ij} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (1)$$

This calculation resulted in a 62*62 connectivity matrix for each subject. All self-connections or negative connections (such as functional autocorrelations) were removed from the networks prior to analysis. The role of negative weights in global network organization may be determined in future studies (Rubinov and Sporns, 2010) but were excluded here.

Due to the instability induced by thresholding procedures and the difficulty of comparing networks constructed for a range of thresholds, we focused on the critical structural network backbone and subjected the networks to two main types of analysis for feature extraction: (i) weighted connectivity matrix and, (ii) a robust tree representation called minimum spanning tree (MST) (Van Steen, 2010; Tewarie et al., 2015; Wang et al., 2008). The former, *weighted connectivity matrix* has been used widely in the literature (Chu et al., 2015; Fraschini et al., 2015) from which we derived two common network metrics: global efficiency and the global and local clustering coefficient (Holland and Leinhardt, 1971; Supekar et al., 2008). (For a comprehensive overview see Rubinov and Sporns, 2010.) The latter, *MST tree representation*, is a relatively novel technique that has not been used widely in biological data, and to our knowledge not applied to infants' raw EEG data.

MST has been used as an alternative method that overcomes thresholding problems. Given a connected, undirected weighted graph, a spanning tree of the graph is a subgraph that is a tree and connects all the vertices together (Stam et al., 2014). A single graph can have many different spanning trees (Fig. 1F). If each edge has a distinct weight then there will be only one, unique minimum spanning tree. Following Gower (Gower, 1996), we define the distance between the time evolution of channels i and j as:

$$d(i,j) = \sqrt{(\rho_{ii} + \rho_{jj} - 2\rho_{ij})} = \sqrt{2(1 - \rho_{ij})} \quad (2)$$

Given the symmetry property of the correlation matrix $\rho_{ij} = \rho_{ji}$ and $\rho_{ii} = 1$ for every channel i, the last equality is given. The MST, constructed from a set of N elements, 62 channels with distances $d(i,j)$ between every pair of elements i and j, is a planar graph with $N - 1$ edges connecting the N elements and minimum total length, i.e. $\sum_{\text{in the MST}} d(i,j)$ is minimum respect to any other tree. No loops exist in the MST. The MST displays only the most important links in each node. Network features extracted from MST were *leaf number* and *tree hierarchy*. For each tree, leaf number is defined as the number of leaves, divided by the maximum number of leaves possible given the size of the tree (Dubbelink et al., 2014). Tree hierarchy is used as indicator of the balance between communication paths and overload prevention (Boersma et al., 2014; Dubbelink et al., 2014). It is defined as:

$$T_H = \frac{L}{2mBC_{\max}} \quad (3)$$

where L is the leaf number; m the number of vertices −1; and $BC_{\max}$ the maximum value of betweenness centrality. If leaf number = 2 (i.e. a line-like topology), and m approaches infinity, tree hierarchy approaches 0. If leaf number = m (i.e. a star-like topology) tree hierarchy approaches 0.5. For leaf numbers between these two extreme situations, tree hierarchy can have higher values (Dubbelink et al., 2014).

3.2. Feature selection

We calculated the following features: (i) global efficiency, (ii) global and local clustering coefficient from weighted networks, and (iii) leaf number, and (iv) tree hierarchy, using MST for six different frequency bands.

The significance of the difference ($p < 0.05$) across the two groups under study (FH+ and FH−) were examined with *t-tests* defined as (t-test: $t(F_{FH+}, F_{FH-})$, p-value). F stands for feature name. Meaningful decreases in efficiency were detected in the $\delta(t(Eff_{FH+}, Eff_{FH-}) = 2.83,\ p = 0.016)$, $\theta(t(Eff_{FH+}, Eff_{FH-}) = 2.62,\ p = 0.023)$, $\alpha_1(t(Eff_{FH+}, Eff_{FH-}) = 2.20,\ p = 0.049)$. Similarly, the global clustering coefficient showed a significant difference in $\delta(t(Clus_{FH+}, Clus_{FH-}) = 3.44,\ p = 0.0055)$, $\theta(t(Clus_{FH+}, Clus_{FH-}) = 2.98,\ p = 0.0125)$, $\alpha 1(t(Clus_{FH+}, Clus_{FH-}) = 3.03,\ p = 0.0113)$ and $\alpha 2(t(Clus_{FH+}, Clus_{FH-}) = 2.89,\ p = 0.0147)$ bands. Global efficiency is inversely proportional to the average path length of whole brain network. Significant increases in normalized path length and reduced normalized clustering, suggests reduced cortical communication capacity early in development for infants at higher risk of LLD. In the topogram shown in Fig. 2(A), channels are highlighted that show significant between-group differences in electrode clusters mapped by fre-

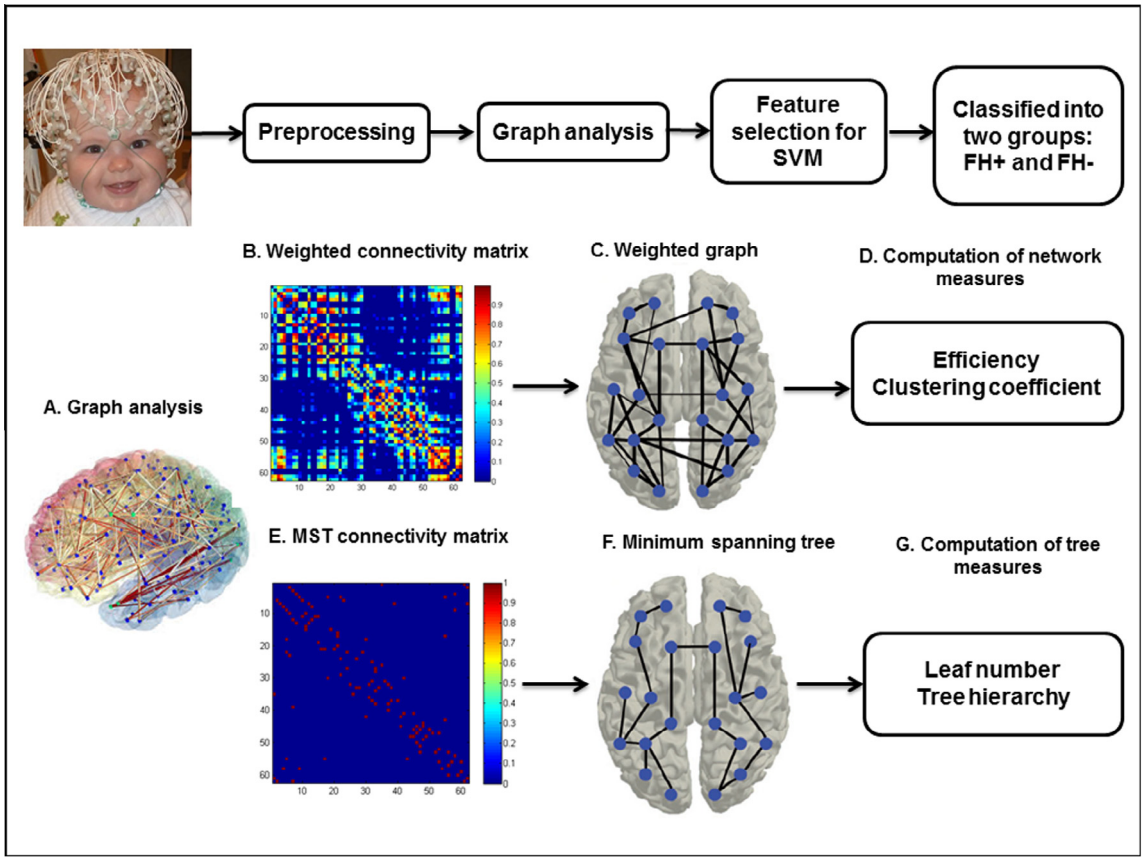

Fig. 1. A schematic overview of infants' EEG data analysis with a focus on the two methods of graph analysis used (A). The first analytic method used was (B) weighted connectivity matrix. For the weighted network analysis a weighted graph (C) was constructed. Network measures computed (D) were efficiency and the clustering coefficient. The second measure was the (E) minimum spanning tree (MST) connectivity matrix. For Tree analysis, a tree representation (F) was constructed. Tree measures computed were: (G) Leaf number and Tree hierarchy.

quency band (FH−, $N = 12$; FH+, $N = 12$). Fig. 2(B and C) depict the grand average FFT for two illustrative channels (31 and 45).

In MST graphs, *leaf number* shows meaningful differences across two groups within a frequency band δt (L_{FH+}, L_{FH-}) = −2.21, $p = 0.042$). Also, tree hierarchy captures significant differences across two groups in one frequency band ($t(TH_{FH+}, TH_{FH-}) = -2.29$, $p = 0.042$). See Fig. 3 for illustrative purposes. Table 1 summarizes the *p*-values of the explained features.

Considering only features that show ($p < 0.05$) helps to ensure that undistinguishable features are removed from the selected best features and thus will not decrease overall performance. About 105 features out of 910 features were selected to classify the two groups. We performed a random shuffling procedure on about 2000 possible feature combinations, and applied the SVM classifier to affirm that the performance of the classifier is not random, and specificity, precision and accuracy of the classifier, using the selected features, lies within the region of ($p < 0.05$).

3.3. Support vector machine (SVM) classification

Support Vector Machine (SVM) is a technique for supervised classification based on the concept of decision planes (Cortes and Vapnik, 1995). In addition to performing linear classification, the SVM method uses training data employing a non-linear classification called the kernel trick and constructs a hyperplane to map groups in *n*-dimensional feature space (n is the number of features). New examples are then mapped into the same space and predicted to belong to a category based on which side of the gap they fall on. Here, we deployed one linear kernel as well as three nonlinear kernel functions including quadratic or polynomial of degree 2, cubic or polynomial of degree 3, and a multilayer perceptron (MLP) function (Cybenko, 1989). The performance of SVM classifiers was assessed using leave-one-out-cross-validation (LOOCV), averaging performance across N sets. LOOCV is a special case of cross-validation where the number of sets equals the number of instances in the dataset. Note that LOOCV is independent of the feature selection technique, confirms the accuracy of the set of features that are selected by the SVM, and provides a feature set that is unbiased by group. Features are used as an input to a support vector machine (SVM) that performed supervised classification, mapping infants into two groups: FH+ infants with the risk of LLD and FH− infants without the risk of LLD. (See Fig. 4 for a schematic overview of the feature selection process and Random Shuffling performance assessment.)

We assessed the classifier's performance using the conventional measures of precision, specificity and accuracy. The FH+ group with a high risk of LLD is designated, positive (P) and the FH− controls are categorized as negative (N). The correct detection or "true classification" of the high-risk condition is known as true positive (TP). Likewise correct classification of the typical low-risk FH− population is true negative (TN). *Precision or Positive predictive value* (*PPV*) measures the proportion of positives, which are correctly identified as such (e.g. the percentage of infants who are correctly identified as having high risk for LLD). *Specificity* is the percent of participants correctly classified as low-risk within all low-risk infants also known as the true negative rate (TNR). These measures along with *accuracy rate* (*ACC*) are given by the following equations:

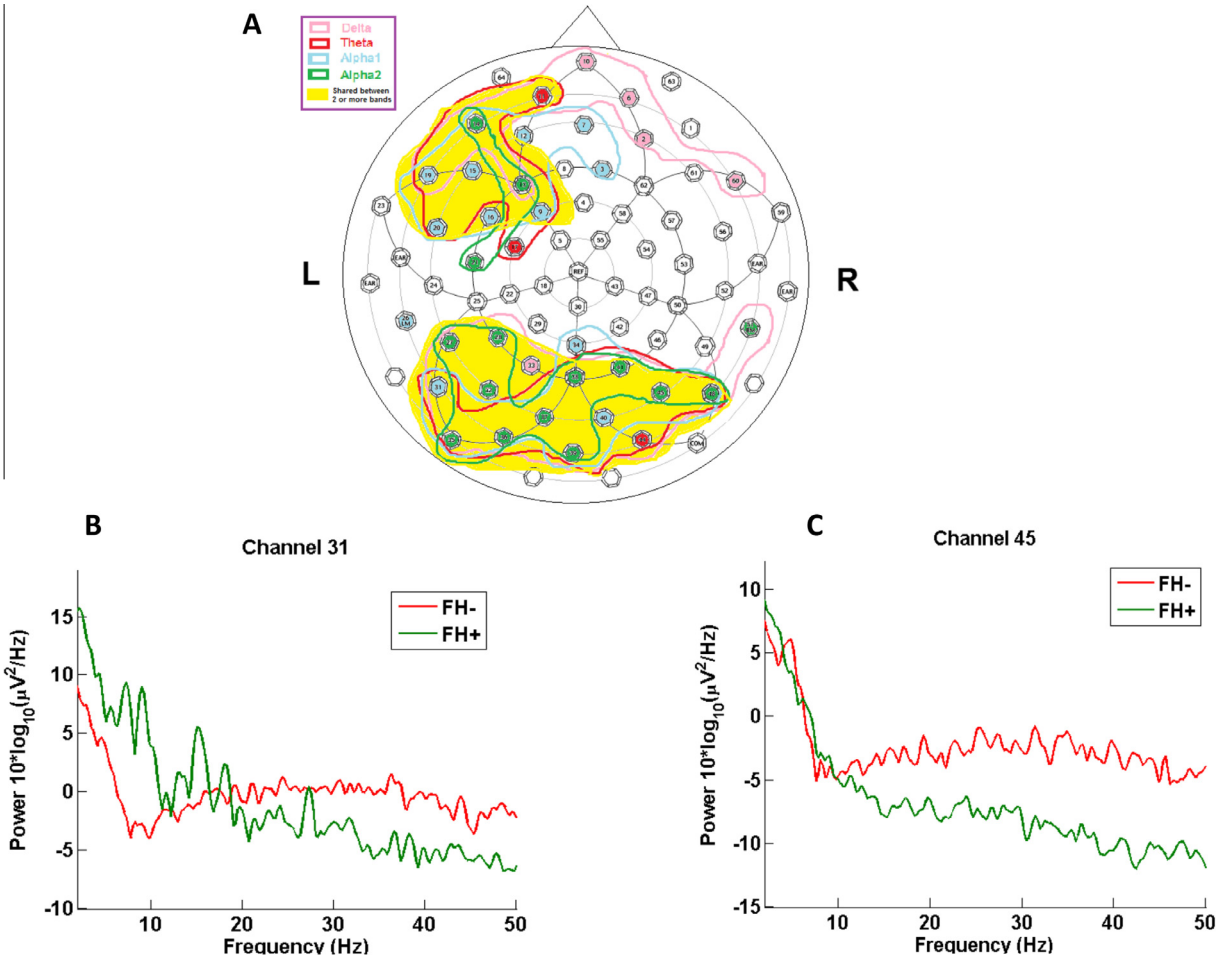

Fig. 2. Between-group differences in electrode clusters mapped by frequency band on a 64-channel EGI template and illustrative FFT plots. Significant group differences in the local clustering coefficient (assessed by *t*-test) were seen between the two infant groups (FH+ and FH) at 6 months-of-age. (A) Channel clusters with $p < 0.05$ were identified within all four bands (Delta, Theta, Alpha1 and Alpha2); cluster patterns that differed by group were concentrated over left frontal and occipital regions. Here, channels are separated by color to show channel cluster membership by frequency: pink for the "delta" band, red for "theta", blue for "alpha1" and green for "alpha2". Regions of between-group differences that were shared across frequencies are highlighted in yellow. In (A), channels are highlighted that show significant differences for the two groups (FH−, $N = 12$; FH+, $N = 12$) by frequency in the local clustering coefficient. The delta band is shown in pink, theta in red, alpha 1 in blue and alpha 2 in green. We have also plotted, in yellow, the channel clusters that share significant differences in two or more frequency bands. The left frontal cluster includes channels [9,11,13,14,15,16,19,20] and the lower occipital cluster includes channels [27,28,31,32,33,35,36,37,38,39,40,41,44,45,48]. Six channels show significant differences across all frequency bands [13, 14, 37, 38, 41, 45]. To illustrate further, (B) shows the grand average FFT by group for Channel 31 with differences in Delta, Theta and Alpha 1 and (C) shows the grand average FFT for Channel 45 with significant group differences across all four frequency bands.

$$\text{Precision or PPV} = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$\text{Specificity or TNR} = \frac{TN}{TN + FP} \times 100 \quad (5)$$

$$\text{Accuracy or ACC} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

Network measures and SVM-related computations were carried out in MATLAB 7.8.0 (R2014; The Mathworks, Natick, MA), as well as with functions available as part of the MATLAB Bioinformatics and Statistics Toolboxes, and in-house MATLAB code. Network features were computed in MATLAB using Brain Connectivity Toolbox (BCT).

To assess the performance of SVM classifiers, we deployed a statistical training test. Training the SVMs within different groups allowed us to test the classifiers' generalization to data not trained on. Each data set can then be tested on each of the trained SVMs, and the group with high risk of LLD classification accuracy would then be the average of these tests.

4. Results

Significant differences were seen in the main features examined both for the global network and for the filtered frequency bands. Table 1 summarizes those data by feature and by network.

The MST results detailed in Table 1 suggest quite different functional connectivity among brain networks as a function of risk status in combination with a non-biased examination of features. Significant increases were seen in normalized path length (i.e. lower efficiency) and reductions in normalized clustering for infants at familial risk for LLD as well as a subset of infants with no such familial risk. In addition, both leaf number and tree hierarchies significantly differed by group. Increases in leaf number in the FH+ group suggests decentralization of early brain networks; in essence, the infants classified as at higher risk for LLD, show networks that are more hierarchical, but inefficient with scattered connections that slow information flow within the network. Although the network is more "leafy", it is more inefficient as captured by the decreases in efficiency and clustering.

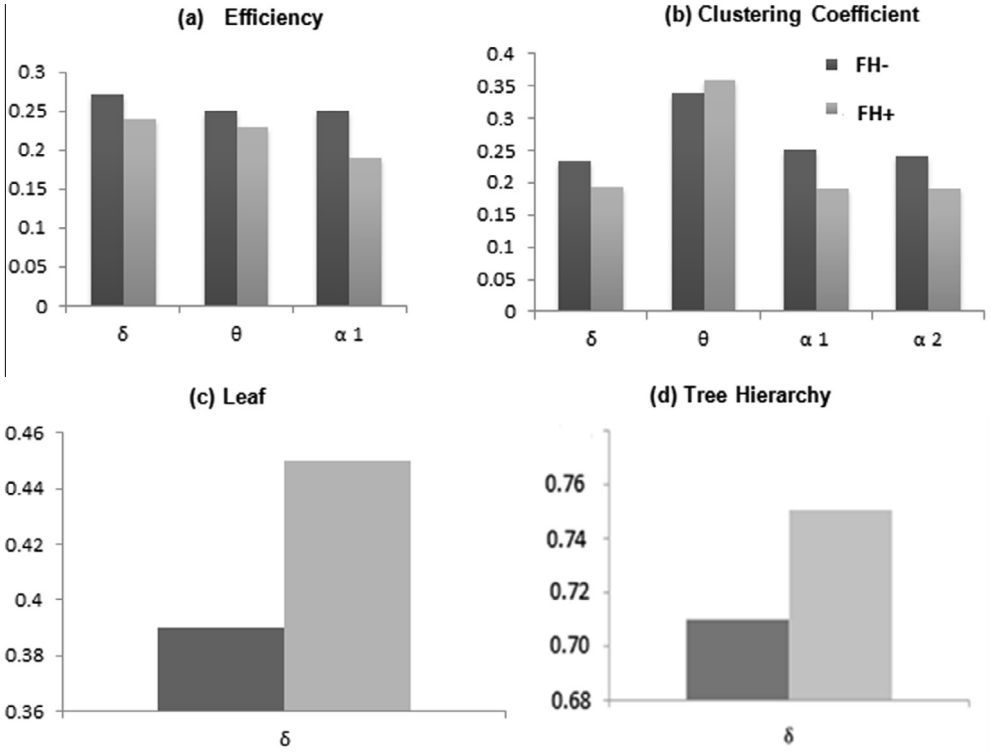

Fig. 3. Significant differences ($p < 0.05$) across two groups of infants for four main frequency bands (x-axis) is shown for each of the four Features: (a) Efficiency shows ($p < 0.05$), and decreases between FH− and FH+ are significant within three of the frequency bands (δ, θ, α_1) with FH+ less efficient, (b) Clustering coefficient is shown ($p < 0.05$) within four frequency bands, the coefficient increases in the θ− band, while it decreases in the other three bands (δ, α_1, α_2) with FH+ showing less clustering than the FH− group; (c) Leaf number is shown ($p < 0.05$) in two frequency bands (δ, γ), and increases for FH+ as compared to FH−. (d) Tree hierarchy is shown in one frequency band (δ), and it increases for the FH− as compared to FH+.

Table 1
Significant differences ($p < 0.05$) are indicated by an asterisk for each of the four main features seen across the two groups (FH+ and FH−) for the Whole Network and for the six networks derived from filtering into the six frequency bands.

	Whole network	δ	θ	α_1	α_2	β	γ
Efficiency	0.35	**0.016***	**0.023***	**0.049***	0.06	0.18	0.47
Clustering coefficient	0.14	**0.0055***	**0.01***	**0.01***	**0.01***	0.21	0.91
Leaf number	0.11	**0.042**	0.61	0.39	0.60	0.43	0.83
Tree hierarchy	0.36	**0.042***	0.87	0.71	0.57	0.31	0.79

The SVM test indicated that cubic kernel was able to identify two classes of subjects: low and high risk of LLD groups with high specificity, precision and accuracy. The results are shown in Table 2. As seen here the Cubic kernel function has the highest performance among the other kernel functions. It is notable that SVM performance without LOOCV is 100% for all performance measures.

To further assess the significance of the specific features classifying the two groups, we ran SVM on folds of 105 features by randomly shuffling all possible features with 2000 iterations and then assessed the performance of the cubic SVM for each fold. Our results indicate that the classification accuracy with the selected features based on significant differences across the two groups were robust and reliable (Table 3). As is clear from the table, specificity, precision and accuracy of the classifier reported in the current analyses lies within the region ($p < 0.05$).

Although at this stage, conclusions based on graph metrics and automatic classification are speculative, and much research remains to be done, these findings suggest reduced (or altered) functional networks and cortical communication capacity quite early in development for infants at higher risk of LLD.

5. Discussion

Over the infancy period, children demonstrate striking increases in language, cognitive and motor skills (Bates et al., 1992; Chugani et al., 1987; Fenson et al., 2007; Moll and Tomasello, 2010; Pascalis et al., 2005,). This developmental stage is also characterized by major changes in forebrain organization including continuing maturation and myelination of temporal and frontal areas and elaboration of extensive cortical and subcortical circuits (e.g. Johnson, 2001; Paus et al., 2001; Pujol et al., 2006). Although many developmental disorders appear to have their earliest roots within the infancy period, it is notoriously difficult to detect these early risk markers and thus accurately predict to later developmental disorders such as LLD.

Accumulating data have shown that measures of rapid auditory processing (RAP) ability can serve as a behavioral "marker" of LLD and can reliably identify infants at highest risk of LLD (Benasich and Leevers, 2002; Choudhury and Benasich, 2011). However, reliable identification of such markers would benefit greatly from more sophisticated and sensitive convergent measures. EEG has

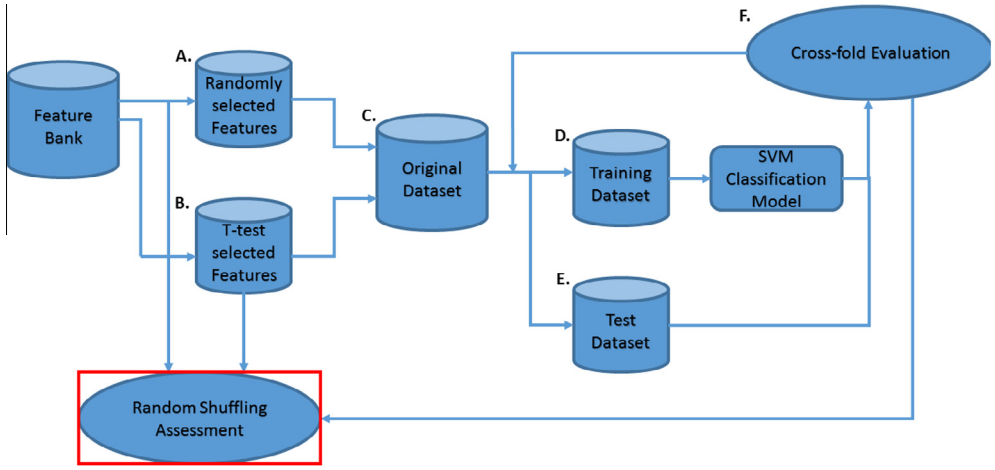

Fig. 4. SVM classification and performance assessment by Random Shuffling. After feature extraction, a subset of "best features" need to be identified. To select appropriate features from the feature bank, we used two approaches: Random selection (A) and selection via *t*-test (B). For both approaches, the original dataset (C) is divided into two sets: Train (D) and Test (E), using cross-validation methodology. An assessment measure is calculated to evaluate each selected feature via the support vector machine classifier. Finally, a random shuffling cross-fold evaluation (F) is performed to ensure that the *t*-test selected features are statistically significant and that these are the "best features" based on the assessment measures.

Table 2
Specificity, precision, and accuracy of the four SVM kernel functions.

Kernel function	Specificity	Precision	Accuracy
Linear	0.7692	0.7500	0.7917
Quadratic	0.6250	0.7500	0.5833
MLP	0.6667	0.5833	0.7083
Cubic	**0.8889**	**0.9167**	**0.7917**

Table 3
Distribution of specificity, precision and accuracy of SVM performance in random shuffling of the features.

	0–20%	20–40%	40–60%	60–80%	80–100%
Specificity	0	0	129	1851	20
Precision	0	86	1068	775	71
Accuracy	0	0	71	1825	102

become an extremely valuable technique for acquiring longitudinal data defining maturational brain trajectories (e.g. Choudhury and Benasich, 2011) and for analyses of developmental power-bands and oscillation data that explore the evolution of dynamic coordination (oscillatory signatures) and synchrony in the developing brain (Gou et al., 2011. Musacchia et al., 2013; Ortiz-Mantilla et al., 2013). Despite these advances, more sophisticated computational approaches to developmental data, including EEG network analysis of infant data, are relatively rare. Such techniques might well add a valuable dimension to the difficult task of early screening, diagnosis and intervention. The ability to examine the conjunction of multiple converging features, something that is very difficult to accomplish in traditional EEG and ERP approaches, seems likely to add significant value to standard clinical approaches. To this end, focusing on network connectivity measures of infants across the first year of life and their utility in identifying early markers of LLD, as well as the design and realization of a robust and reliable automatic classification of infant risk should be a priority.

Automatic classifiers might well aid in achieving even earlier detection of neural precursors of developmental disorders, thus providing the ability to differentiate infants at highest risk long before actual clinical diagnosis is possible (Bosl et al., 2011; Vourkas et al., (2014)), an important goal in the field. Our analysis using an SVM classifier on features extracted from a graph analysis, which included standard measures (i.e. efficiency and global and local clustering coefficient), and results from a new tree-based analysis, MST (i.e. leaf number and tree hierarchy) suggest that infants in families with a history of LLD have quite different EEG complexity patterns from infants without such familial risk. Specifically, infants were classified with about 80% accuracy into FH+ and FH− groups at 6-months-of age with 89% specificity and with precision of 92%. Accuracy represents the percentage of accurately classified subject (FH+ or FH−) compared to the total number of subjects. Precision or PPV (Positive predictive value) captures the accuracy of the classifier method. The smaller the "false positive" (FP) rate, the closer precision is to 1. When the goal is to correctly classify subjects as "high-risk" and we identify ALL subjects who actually are "at risk" of the disorder, precision would be 1. Specificity is the rate of TN or "True Negatives" to the total number of negatives, i.e. number of subjects classified as True "low-risk" to the total number of "low risk". Specificity would be 1 when the subjects with "high-risk" are not classified as "low-risk".

Although the present analysis infrequently classifies subjects with low risk of LLD as TP or "True Positives", these "mis classifications" may contain important information regarding future risk. It would be necessary to follow these infants through 5 or 6 years-of-age to make a definitive statement as to whether using an automatic machine classifier approach substantially improves early prediction of LLDs. However, given the dearth of diagnostic tools available at this point in time, designing and training a machine to detect infants at highest risk of LLD, using a combination of features from network analysis, may well enable more accurate predictions as to whether any particular child, with or without a familial history, will actually develop a LLD. As noted above, about 30–60% of infants at familial risk go on to develop a LLD (Flax et al., 2003; Tomblin, 1989), and across our previous studies 8–10% of children with NO identified familial risk (i.e. FH− Control infants) later develop LLDs (Benasich and Leevers, 2002). Thus you can see that family history alone does not constitute a "gold standard". The high level of classification accuracy shown here for this machine learning approach suggests that such methodologies can significantly increase our ability to differentiate those infants at highest risk for LLD, thus allowing the earliest possible intervention and remediation (e.g. Benasich et al., 2014).

SVM classifiers have been applied to facilitate detection of other disorders including children at familial risk for autistic spectrum disorders (e.g. Bosl et al., 2011; Ebrahimi et al., 2008; Sacchet

et al., 2015) and to study brain networks in patients with Alzheimer's disease (Khazaee et al., 2015); however, we are not aware of any research that employs the SVM classifier on features extracted from infant EEG using the MST technique. Vourkas et al. (2014) have deployed MST analyses to evaluate function-related electrocortical reactivity over a wide range of EEG frequencies in school-age children. Using minimum spanning tree (MST), EEG signals from children with math difficulties (MD) and their typically achieving controls (NI) were analyzed. Although no significant differences were found between the groups, a subset of MST parameters correlated with individual performance on psychometric mathematical exams. Thus their findings lend support to the potential utility of MST analysis. In the present study the MST analysis highlighted important group differences in leaf number and in tree hierarchy, suggesting less well organized brain networks and reduced cortical communication capacity early in development for infants at familial risk for LLD. Thus this seems to be a valuable technique to add to a machine learning approach to EEG risk analysis.

Automatic classifiers may be able to detect early risk markers that cannot easily be identified by examination of behavior and averaged EEGs, given that machine learning techniques can detect widely distributed but potentially less robust information captured within a subset of specific features. Even if such a classifier does not precisely define which type of language-based disorder the infant might develop in the future, it adds convergent information that may allow proactive intervention or treatment of infants well before the disorder emerges.

Assessment of SVM performance in these analyses indicates that our classifier is sensitive, robust and reliable; however, it does not provide an exact detection method. In combination with other diagnostic techniques, however, as part of an overall assessment, this technique may yield a clinically useful risk biomarker. The novel MST method we applied to infants' EEG data is also a promising network analysis technique and hopefully will enable deeper understanding of the relationship between neurophysiological processes as measured by EEG and later language outcome. The MST method might serve to enhance feature detection in future longitudinal analyses of data from this cohort by constructing a unique tree that represents the connections between nodes in the network as a function of risk group. The MST technique essentially overcomes the problems induced by thresholding that researchers face using network analysis. It is, in essence, a technique that facilitates network feature selection. "Efficiency" of a network refers to how efficiently information is exchanged and communicated across the network and in the MST analysis; this feature was shown to distinguish the two groups from each other. Efficiency is the most common measure used to index network disruption due to a disease or disorder (Boersma et al., 2014; Rubinov and Sporns, 2010).

Study Limitations: A few caveats should be raised. First, one might argue that given the millisecond temporal resolution of EEG data, these types of network analyses may not prove as reliable as has been shown in fMRI studies. However, the robust classifier results reported above suggest that is probably not the case. Further, EEG assessment has additional strengths as a research and diagnostic technique, particularly in infants, as it is less invasive, less prone to movement artifact and inexpensive as compared to fMRI. Another possible criticism regarding network analysis of EEG data is that the feature selection itself may be unreliable. However, the results of our random shuffling procedures including all possible features, as well as the "leave-one-out-cross-validation", that was run on the data set and then generalized to an independent data set suggests that this is not a valid concern. It is also important to highlight the fact that this is a relatively small sample and thus the generalizability of these findings is limited. Thus, we plan to apply these machine learning techniques to larger developmental data sets and importantly to longitudinal data sets. However, the primary aim of this study was to illustrate the utility of SVM techniques to improve early identification of those children most at risk for LLD and ultimately facilitate remediation at those developmental time points when the brain is most plastic and amenable to intervention techniques.

6. Conclusions

Overall, these results are encouraging and suggest that the objective of using machine learning techniques (e.g. SVM) and MST analyses to discriminate between groups of infants at differing levels of risk for LLD and then applying cross validation to assess the resultant classification is a viable technique. Although this is a promising analytic technique, it is certainly the case that further research and development is critical and the next steps should incorporate analysis of prospective longitudinal samples.

We hope that the present findings will provide an impetus to expand the scope of clinical screening in the future with greater dependence on EEG-based screening assessment using a multi-feature approach. Future studies will hopefully focus on trajectory analyses of prospective longitudinal data from infant siblings of children with disorders such as LLD, autism spectrum disorder and attention-deficit hyperactive disorder using spontaneous EEG and MEG as well as event-related EEG, employing machine learning procedures similar to those described here.

In conclusion, we accurately classified infants into two groups at low and high risk of LLD using unbiased features derived from network analysis. Our results are promising and should motivate examination of similar methods that will support more detailed risk assessment in infancy and identification of early, possibly specific, biomarkers for a range of developmental disorders (Bosl et al., 2011). Further research is needed to assess the potential of machine learning methods using averaged ERPs as well as for the spontaneous infant EEG used here. Nonetheless, classification methods (e.g. SVM) and MST, in combination with cross validation, hold the promise of increasing the discriminative power of MEG, EEG and EEG/ERP measurements, reducing problems associated with using large series of univariate group comparisons, and should aid in the quest to distinguish risk or clinical groups from typically-developing control groups.

Acknowledgments

Funding for this research was provided by the Elizabeth H. Solomon Center for Neurodevelopmental Research with additional funding from a National Science Foundation (NSF) Grant SMA-1041755 to AAB as part of the Temporal Dynamics of Learning Center (TDLC), an NSF Science of Learning Center. MZ would like to thank the Iranian Cognitive Sciences and Technologies Council for their grant support to alumni abroad. We particularly want to thank the families who volunteered their time to participate in these studies and give special thanks to all research assistants, psychologists, and speech-language pathologists who helped with infant testing and data collection including N. Choudhury, T. Realpe-Bonilla and C. Roesler. MZ and ZR acknowledge A. Katanforoush for reviewing the manuscript and for helpful advice.

Conflict of interest: The authors declare that they have no conflict of interest.

References

Aslin R. Discrimination of frequency transitions by human infants. J Acoust Soc Am 1989;86:582–90.

Barttfeld P, Wicker B, Cukier S, Navarta S, Lew S, Sigman M. A big-world network in ASD: dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. Neuropsychologia 2011;49:254–63.
Bates E, Thal D, Janowsky JS. Early language development and its neural correlates. second ed. In: Rapin I, Segalowitz S, editors. Handbook of Neuropsychology 1992;vol. 7. Amsterdam: Elsevier; 1992. p. 69–110.
Benasich AA, Choudhury N. Timing, information processing and efficacy: early factors that impact childhood language trajectories. In: Benasich A, Fitch R, editors. Developmental Dyslexia: Early Precursors, Neurobehavioral Markers and Biological Substrates (The Extraordinary Brain Series). Baltimore, MD: Brookes Publishing Co, USA; 2012. p. 99–118.
Benasich AA, Leevers H. Processing of rapidly presented auditory cues in infancy: implications for later language development. Prog Inf Res 2002;vol. 3. p. 245–88.
Benasich AA, Tallal P. Infant discrimination of rapid auditory cues predicts later language impairment. Behav Brain Res 2002;136:31–49.
Benasich AA, Choudhury N, Friedman JT, Realpe Bonilla T, Chojnowska C, Gou Z. Infants as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. Neuropsychologia 2006;44:396–411.
Benasich AA, Gou Z, Choudhury N, Harris KD. Early cognitive and language skills are linked to resting frontal gamma power across the first three years. Behav Brain Res 2008;95:215–22.
Benasich AA, Choudhury N, Realpe-Bonilla T, Roesler C. Plasticity in developing brain: active auditory exposure impacts prelinguistic acoustic mapping. J Neurosci 2014;34:13349–63.
Boersma M, Smit D, Boomsma D, Geus E, Delemarre-van de Waal H, Stam C. Growing trees in child brains: graph theoretical analysis of EEG derived minimum spanning tree in 5 and 7 year old children reflects brain maturation. Brain Connect 2014;3:50–60.
Bosl W, Tierney A, Tager-Flusberg H, Nelson C. EEG complexity as a biomarker for autism spectrum disorder risk. BMC Med 2011;9:18.
Bush G. Attention-deficit/hyperactivity disorder and attention networks. Neuropsychopharmacology 2010;35:278–300.
Casey M, Cohen M, Schuerholz L, Singer H, Denckla M. Language-based cognitive functioning in parents of offspring with ADHD comorbid for Tourette syndrome or learning disabilities. Dev Neuropsychol 2000;17:85–110.
Choudhury N, Benasich AA. Familial aggregation in a sample of infants born into families with a history of language-based learning impairments. J Speech Lang. Hear Res 2003;46:261–72.
Choudhury N, Benasich AA. Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year language and cognitive abilities. Clin Neurophysiol 2011;122:320–38.
Choudhury N, Leppänen PHT, Leevers H, Benasich AA. Assessing rapid auditory processing in family history and control infants: evidence from converging paradigms. Dev Sci 2007;102:213–36.
Chu CJ, Tanaka N, Diaz J, Edlow BL, Wu O, Hämäläinen M, Stufflebeam S, Cash SS, Kramer MA. EEG functional connectivity is partially predicted by underlying white matter connectivity. NeuroImage 2015;108:23–33.
Chugani H, Phelps M, Mazziotta J. Positron emission tomography study of human brain functional development. Ann Neurol 1987;22:487–843.
Cortes C, Vapnik V. Mach Learn 1995;20:273–97.
Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signal 1989;2:303–14.
Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. J Neurosci Methods 2004;134:9–21.
Dubbelink KTO, Hillebrand A, Stoffers D, Deijen JB, Twisk JW, Stam CJ, Berendse HW. Disrupted brain network topology in Parkinson's disease: a longitudinal magnetoencephalography study. Brain 2014;137:197–207.
Ebrahimi F, Mikaeili M, Estrada E, Nazeran H. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. Engineering in Medicine and Biology Society, 30th Conf Proc IEEE, 2008. pp. 1151-1154.
Eilers R, Morse P, Gavin W, Oller D. Discrimination of voice onset time in infancy. JASA 1981;16:186–97.
Elison J, Wolff J, Heimer J, Paterson S, Gu H, Hazlett H, Styner M, Gerig G, Piven J. Frontolimbic neural circuitry at 6 months predicts individual differences in joint attention at 9 months. Dev Sci 2013;70:955–65.
Fenson L, Marchman V, Thal D, Dale P, Reznick S, Bates E. MacArthur-Bates Communicative Development Inventories: User's guide and technical manual. Baltimore, MD: Paul H. Brookes Publishing Company; 2007.
Flax J, Realpe-Bonilla T, Hirsch L, Brzustowicz L, Bartlett C, Tallal P. Specific language impairment in families: evidence for co-occurrence with reading impairments. J Speech Lang Hear Res 2003;46:530–43.
Fraschini M, Hillebrand A, Demuru M, Didaci L, Marcialis G. An EEG-based biometric system using eigenvector centrality in resting state brain networks. IEEE Signal Process Lett 2015;22:666–70.
Gou Z, Choudhury N, Benasich AA. Resting frontal gamma power at 16, 24 and 36 months predicts individual differences in language and cognition at 4 and 5 years. Behav Brain Res 2011;220:263–70.
Gower J. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 1996;51:325–38.
Heim S, Benasich AA. Reduced sensory oscillatory activity during rapid auditory processing as a correlate of language-learning impairment. J Neurolinguistics 2011;24(5):539–55.
Holland PW, Leinhardt S. Transitivity in structural models of small groups. Comp Group Stud 1971;2:107–24.
Johnson M. Functional brain development in humans. Nat. Rev. Neurosci 2001;2:475–83.
Khazaee A, Ebrahimzadeh A, Babajani-Feremi A. Identifying patients with Alzheimer's disease using resting state fMRI and graph theory. Clin Neurophysiol 2015;126:2132–41.
Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos Trans R Soc Lond B Biol Sci 2008;363:979–1000.
Leonard C, Low P, Jonczak E, Schmutz K, Siegel L, Beaulieu C. Anatomy, processing speed and reading in children. Dev Neuropsychol 2011;36:828–46.
Lewis J, Elman J. Growth-related neural reorganization and the autism phenotype: a test of the hypothesis that altered brain growth leads to altered connectivity. Dev Sci 2008;11:135–55.
Maitre N, Lambert W, Aschner J, Key A. Cortical speech sound differentiation in the neonatal intensive care unit predicts cognitive and language development in the first 2 years of life. Dev Med Child Neurol 2013;55:834–9.
Moll H, Tomasello M. Infant cognition. Curr Biol 2010;20:R872–5.
Musacchia G, Choudhury NA, Ortiz-Mantilla S, Realp-Bonilla T, Roesler C, Benasich AA. Oscillatory support for rapid frequency change processing in infants. Neuropsychologia 2013;51:2812–24.
Ortiz-Mantilla S, Hämäläinen JA, Musacchia G, Benasich AA. Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. J Neurosci 2013;33:18746–54.
Pascalis O, Scott L, Kelly D, Shannon R, Nicholson E, Coleman M, Nelson C. Plasticity of face processing in infancy. PNAS 2005;102:5297–300.
Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A. Maturation of white matter in the human brain: a review of magnetic resonance studies. Brain Res Bull 2001;54:255–66.
Pruett JR, Kandala S, Hoertel S, Snyder AZ, Elison JT, Nishino T, Network I. Accurate age classification of 6 and 12 month-old infants based on resting-state functional connectivity magnetic resonance imaging data. Dev Cogn Neurosci 2015;12:123–33.
Pujol J, Soriano-Mas C, Ortiz H, Sebastian-Galles N, Losilla JM, Deus J. Myelination of language-related areas in the developing brain. Neurology 2006;66:339–43.
Riaz F, Hassan A, Rehman S, Niazi IK, Dremstrup K. EMD based temporal and spectral features for the classification of EEG signals using supervised learning. IEEE Trans Neural Syst Rehabil Eng 2013;2:107–24.
Rodgers J, Nicewander W. Thirteen ways to look at the correlation coefficient. J Am Stat Assoc 1988;42:59–66.
Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 2010;52:1059–69.
Sacchet M, Prasad G, Foland-Ross L, Thompson P, Gotlib IH. Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. Front Psychiatry 2015;18:6–21.
Schiavone G, Linkenkaer-Hansen K, Maurits NM, Plakas A, Maassen BA, Mansvelder HD, van Zuijen TL. Preliteracy signatures of poor-reading abilities in resting-state EEG. Front Hum Neurosci 2014;8:735.
Stahl D, Pickles A, Elsabbagh M, Johnson M. Novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism. Dev Neuropsychol 2012;37:274–98.
Stam CJ, Tewarie P, Van Dellen E, van Straaten EC, Hillebrand A, Van Mieghem P. The trees and the forest: characterization of complex brain networks with minimum spanning trees. Int J Psychophysiol 2014;92:129–38.
Stets M, Stahl D, Reid VM. A meta-analysis investigating factors underlying attrition rates in infant ERP studies. Dev Neuropsychol 2012;37:226–52.
Supekar K, Menon V, Rubin D, Musen M, Greicius M. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol 2008;4: e1000100.
Tallal P. Improving language and literacy is a matter of time. Nat Rev Neurosci 2004;5:721–8.
Tewarie P, Dellen E, Hillebrand A, Stam C. The minimum spanning tree: an unbiased method for brain network analysis. NeuroImage 2015;104:177–88.
Tomblin JB. Familial concentration of developmental language impairment. J Speech and Hear Disord 1989;54:267–95.
Van Steen M. Graph Theory and Complex Networks: An introduction. Amsterdam: Maarten van Steen; 2010.
Vourkas M, Karakonstantaki E, Simos PG, Tsirka V, Antonakakis M, Vamvoukas M, Stam C, Dimitriadis S, Micheloyannis S. Simple and difficult mathematics in children: a minimum spanning tree EEG network analysis. Neurosci Lett 2014;576:28–33.
Wang H, Hernandez J, Van M. Betweenness centrality in a weighted network. Phys Rev E 2008;77:2.
Werker J, Tees R. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. Dev Psychobiol 2005;46:233–51.
Westwood P. Learning and Learning Difficulties: Approaches to Teaching and Assessment. David Fulton Publishing Team, Routledge; 2004.
Whitehouse AJ, Barry JG, Bishop DV. Further defining the language impairment of autism: is there a specific language impairment subtype? J Commun Disord 2008;41:319–36.
Wong PC, Ettlinger M, Zheng J. Linguistic grammar learning and DRD2-TAQ-IA polymorphism. PLoS One 2013;8:e64983.