A Machine Learning Approach for Distinguishing Age of Infants Using Auditory Evoked Potentials

Maryam Ravan[a], James P. Reilly[a], Laurel J. Trainor[b], and Ahmad Khodayari-Rostamabad[a]

[a]Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada
[b]Department of Psychology, Neuroscience and Behavior, McMaster University, Hamilton, ON, Canada

Corresponding author:
Maryam Ravan
Address: 1280 Main Street West, Hamilton, Ontario, L8S 4K1
Tel: +1-905-515-9360
Email: mravan@ece.mcmaster.ca

***Keyword*:** Age group determination; machine learning; electroencephalogram (EEG); event related potentials (ERPs), feature extraction, classification; wavelet coefficients.

Acknowledgments

The Natural Science and Engineering Research Council of Canada (NSERC) has funded a large portion of this research through its Discovery Grants program, and also through a Co-Operative Research and Development (CRD) grant, in conjunction with Intratech Inline Inspection Services (I3SL) Ltd., Mississauga, Ontario.

Highlights

- This study demonstrates that machine learning algorithms can be used to classify individual subjects by age group (6-month old, 12-month old, and adult) based on auditory event-related potentials (ERPs).
- The method is unique in that it assumes no *a priori* structure, such as the composition of ERP components, on the ERP signal.
- The proposed algorithm is capable of classifying the age of single subjects, enabling clinical application where abnormal neural development may be indicated when the chronological age of the subject differs significantly from the age determined by the proposed method.

Abstract

Objective: To develop a high performance machine learning (ML) approach for predicting the age and consequently the state of brain development of infants, based on their event related potentials (ERPs) in response to an auditory stimulus.

Methods: The ERP responses of twenty-nine 6-month-olds, nineteen 12-month-olds and ten adults to an auditory stimulus were derived from electroencephalogram (EEG) recordings. The most relevant wavelet coefficients corresponding to the first- and second-order moment sequences of the ERP signals were then identified using a feature selection scheme that made no *a priori* assumptions about the features of interest. These features are then fed into a classifier for determination of age group.

Results: We verified that ERP data could yield features that discriminate the age group of individual subjects with high reliability. A low dimensional representation of the selected feature vectors show significant clustering behavior corresponding to the subject age group. The performance of the proposed age group prediction scheme was evaluated using the leave-one-out cross validation method and found to exceed 90% accuracy.

Conclusions: This study indicates that ERP responses to an acoustic stimulus can be used to predict the age and consequently the state of brain development of infants.

Significance: This study is of fundamental scientific significance in demonstrating that a machine classification algorithm with no *a priori* assumptions can classify ERP responses according to age and with further work, potentially provide useful clues in the understanding of the development of the human brain. A potential clinical use for the proposed methodology is the identification of developmental delay: an abnormal condition may be suspected if the age estimated by the proposed technique is significantly less than the chronological age of the subject.

1. Introduction

Electroencephalography (EEG) has become a prominent method for studying auditory perception in infants (e.g., de Haan, 2007; Trainor, 2008). The EEG is a non-invasive procedure that allows an experimenter to record brain responses from multiple electrodes on the scalp. The event-related potential (ERP) is obtained from the EEG and is any stereotyped electrophysiological response to an internal or external stimulus.

ERPs can also be used to examine auditory perception in very young infants since they do not require any overt behavioural response or direct attention (de Boer et al., 2007; Kropotov et al., 1995). They are one of the few methods that can easily and safely be used to study the rapid development of the brain in infants, and have led to exciting discoveries about human brain functioning and the neural basis of cognition. The evoked response from an auditory stimulus consists of a series of positive and negative deflections (components) in the recorded EEG signal that occur at characteristic times with respect to the time of occurrence of the stimulus. Responses to the repeated presentation of the same stimulus are typically averaged together in order to reduce noise. The resulting waveforms reflect the underlying neural activity from processing the stimulus.

ERPs consist of many different components. The components present and their latencies and morphologies change greatly with development (Taylor and Baldeweg, 2002; Trainor, 2008). Furthermore, at a particular developmental stage, an ERP component may be affected by the type of auditory stimulus presented, the rate of presentation, and state of the subject (asleep, awake, alert,

attending, etc.) to a much larger degree than in adulthood. Thus, the determination of the subject's age based solely on an analysis of the ERP components is a complex process (e.g., Ceponiene et al., 2002; Choudhury and Benasich, 2010; de Haan, 2007; He et al., 2007, 2009a,b ; He and Trainor, 2009; Kushnerenko et al., 2002a,b; Morr et al., 2002; Trainor et al., 2001, 2003). One component of the ERP that can be elicited across a wide age range by auditory stimulation is the mismatch negativity (MMN) response (e.g., Näätänen et al., 2007; Picton et al., 2000). MMN is elicited when a repeating auditory stimulus is occasionally altered in some manner. Even before the adult-like MMN is elicited, infants produce mismatch responses (MMRs) to a change in stimulus. For example, He et al. (2007) found that 2-month-olds generated a slow positive MMR in response to occasional pitch changes in a repeating piano tone, whereas 3- and 4-month-old infants generated negative MMRs similar to the adult MMN in response to this simple pitch change. In another study, Tew et al. (2009) used the same method as He et al. (2007) to examine whether young infants could detect changes in the relative pitch of a melody in transposition. In this study the stimulus consisted of a 4-note melody that was transposed (starting on a different note on different trials) to related keys from trial to trial. Occasionally the last note was changed by a semitone ($1/12^{th}$ octave). This study also demonstrated different MMRs with age, but for this more complex stimulus, 6-month-old infants produced positive slow MMR and adults faster MMN. Thus, these previous studies have suggested that a conclusive determination of age based solely on an analysis of the ERP components is complicated by many factors, including the fact that the ERP patterns which discriminate age vary according to the complexity of the stimulus.

Furthermore, because infants will only remain awake and content, and therefore testable, for a short period of time, in both of the described developmental MMR studies the differences across age in the MMRs were not discernible in individual infants. Therefore, averaging over all ERP trials of all subjects in each age group was performed to improve the signal-to-noise ratio so that the difference in the MMRs across age could be observed. This averaging procedure determines only the aggregate behaviour of the entire group, but for clinical purposes, reliable categorization of maturation is needed in individuals. In this paper, we introduce a new approach that enables the classification into age group for single subjects

that, unlike previous studies, is not explicitly based on an ERP model and hence incorporates no *a priori* assumptions about the ERP components present. The proposed approach potentially exploits all the relevant information present in the ERP signal, whereas determining age by characterizing only the ERP components may result in some information present in the ERP being lost. For example, as we see later, the proposed method directly incorporates features relating to cross-couplings between electrodes, whereas previous methods do not explicitly use this information. ~~Thus, imposing an ERP model can be too restrictive. In this vein, we have developed a high performance machine learning method to classify the developmental age of each individual subject. The ability of the proposed method to classify individual subjects is crucial to its use in a clinical application.~~

The fact that the proposed methodology can classify individual subjects enables several important clinical applications in psychology, psychiatry and neurology, such as the diagnosis of brain injuries, disorders in the central nervous system, or delayed neurological development. The present approach differs from previous clinical approaches (e.g., Friederich and Friederici, 2006; Guttorm et.al., 2001) in that it does not select *a priori* aspects of the ERP to examine. In the present case, since the method applies to determination of age, an important question such as abnormal development is indicated when the chronological age of an infant is considerably greater than the age determined by the classification procedure. Additionally, from a theoretical perspective, the features selected by the machine learning process that are highly indicative of age could potentially give us important clues in the understanding of infant brain development.

The Machine Learning field evolved from the broader field of Artificial Intelligence, which aims to mimic intelligent abilities of humans by machines. One of the goals of machine learning is to automatically extract salient features from a given data set that are most statistically dependent upon the outcome variable, which in this case is the age group of the subject. These features are then applied to analyze new cases. Hence, learning is not only a question of remembering (or learning) but also of generalization to unseen cases.

Machine learning methods have been used previously in the analysis of EEG signals for various medical applications. For example, Greene et al. (2007) developed a method for the detection of seizures in infants. The system uses a linear discriminant classifier to classify ictal and interictal epochs of one-minute duration. Also, Ghosh-Dastidar et al. (2008) used the cosine 'radial basis function neural network' (RBFNN) model to classify the EEG of normal subjects versus epileptic subjects during ictal and interictal periods. In another study, Krajča1 et al. (2007) developed a new method for automatic sleep stage detection in neonates, based on time profile processing using a fuzzy c-means algorithm. Khodayari-Rostamabad et al. (2010) used machine learning methods to predict the response of schizophrenic subjects to the potentially harmful but effective anti-psychotic drug clozapine. In the present paper, we show that a machine learning method can classify ERP data by age.

2. Methods

2.1 The EEG data used for analysis

The objective of the current classification problem is the assignment of subjects to one of the three predetermined age groups, corresponding to 6- and 12-month-old infants, and adults. These age groups are of interest because phoneme processing in speech (e.g., Curtin and Werker, 2007; Kuhl, 2008) and rhythm processing in music (e.g., Hannon and Trainor, 2007) become specialized between 6 and 12 months of age for the particular language and musical system the infant is exposed to. A total of 58 healthy subjects consisting of twenty-nine 6-month-olds, (15 male, 14 female; mean age = 6 months and 4 days, SD = 28 days) nineteen 12-month-olds (9 male, 10 female; mean age = 11 months and 18 days, SD = 25.7 days), and ten adults (2 male, 8 female; mean age = 24, SD = 2.86 years) with no known hearing deficits were included in the present study. Infants were recruited as part of the McMaster Infant database from hospitals in the Hamilton, Ontario, Canada area. It should be noted that the machine learning algorithm is quite robust for different sample sizes between groups.

The stimulus files were 300 ms grand piano timbre tones created through MIDI and the synthesizer program, Creative SB (Creative Technology Ltd., CA). The sound intensity of each tone was normalized

using Adobe Audition (Adobe Systems Incorporated, San Jose, CA). The tones were then combined to produce a standard short 4-note (1200 ms) melody consisting of two rising intervals followed by a falling interval (E F G C) using MATLAB (The MathWorks, Inc., Natick, MA). The melodies were presented in 20 different transpositions, with starting notes ranging between G3 (294 Hz) and D5 (784 Hz). Each successive transposition was always to a related key (i.e., up or down a perfect 5th, 7/12 octave or a perfect 4th, 5/12 octave) from the current key, in a randomized order. Occasional deviant trials contained a wrong last note, but these were not analyzed in the present paper. Melodies were separated by a 700 ms inter-stimulus interval (ISI). The 200 ms prior to melody onset was used as the pre-stimulus baseline reference.

The stimuli were played using E-prime 1.2 software (Psychology Software Tools, Inc., Pittsburgh, PA) from a Dell OptiPlex280 computer through a speaker (WestSun Jason Sound JS1P63, Mississauga, ON) which was located approximately one meter in front of the subject, at a level of 70 dB(A). The adults were instructed to sit quietly and as still as possible for the duration on the experiment, and infants were kept as still as possible. A silent movie was played to keep the subjects happy and still. Attention to the auditory stimuli was not necessary to elicit the desired EEG samples.

The EEGs were recorded with a sampling frequency of 1000 Hz using HydroCel GSN (HCGSN) sensor nets (Electrical Geodesics, Inc., Eugene, OR) with 128 electrodes. The data were then filtered continuously offline using band-pass filter settings of 0.5-20 Hz by first passing the data through a Blackman-weighted low-pass FIR filter of length 195 with a cut-off frequency of 20 Hz and then passing the resulting data through a second Blackman-weighted high-pass FIR filter of length 7683 with a cut-off frequency of 0.5 Hz. Both of these filters have a very flat frequency response and linear phase in the pass-band, thus minimizing distortion in the output signal.

The data were then down-sampled offline at f_s=250 Hz and segmented into epochs of 1900 msec duration (200 msec pre-stimulus baseline, 1200 msec stimulus, and 500 msec post-stimulus interval). Using a sampling frequency of f_s=250Hz, each trial has $N_e = 475$ samples. The entire experiment

contained $M_s = 480$ standard and $M_d = 120$ deviant trials on each subject. Therefore the total experiment length was 19 min.

EEG artifacts were then removed using the artifact-blocking (AB) algorithm (Mourad et al., 2007), a technique that enables artifact removal without eliminating any trial. Only standard trials were analyzed. The individual trials from each electrode were all averaged together and re-referenced by subtracting from the averaged signal obtained over all electrodes. The electrodes were then divided into ten regions, consisting of frontal right and left (8 electrodes each), central right and left (10 electrodes each), parietal right and left (9 electrodes each), occipital right and left (9 electrodes each) and temporal right and left (9 electrodes each) for statistical analysis, as shown in Fig. 1. ERP responses from the electrodes in each region were averaged together. Certain electrodes were not included: 10 electrodes on the midline were excluded so that the ERP responses could be compared across hemispheres, 10 electrodes were excluded from the front of the cap to reduce artifacts due to the eye movements, and 10 electrodes were removed from edge of the cap to reduce the myoelectric effects of neck movements.

2.2 An overview of the machine learning procedure

We now present a brief summary of the machine learning process used for the determination of age group. A somewhat more detailed explanation of machine learning in the clinical context is available in (Khodayari-Rostamabad et al., 2010). A necessary component of this process is the existence of a set of training patterns (subjects). In our case, this set consists of the ERP data of all ten regions in addition to the age group designation (target variables) $y_i \in C,\ i = 1,\ldots M_t$ corresponding to each subject, where $C = \{1, 2, ..., N_c\}$, N_c is the number of classes and M_t is the number of training patterns. In this study, $N_c = 3$ and the corresponding age groups are 6-month-olds, 12-month-olds, and adults, respectively. The value of M_t is 58.

We first compute candidate *features* from the ERP data. For this study, the set of candidate features consists of a discrete wavelet decomposition (DWT) of first- and second-order cumulant functions extracted from the ERP data, as described in more detail in subsection 2.3. The number N_f of such

candidate features can be quite large. The result of the feature extraction process is a set of M_t vectors $\tilde{\mathbf{x}}_i \in \mathrm{R}^{N_f}, i = 1,\ldots,M_t$. After extracting candidate features, the next step is *feature selection,* which will be described in more detail in subsection 2.4. This procedure is critical to the performance of the resulting classifier or predictor. Feature selection is an ongoing topic of research in the machine learning community. Typically, only a relatively small number of the candidate features bear any significant statistical relationship with the target variables. We therefore select only those features that share the strongest statistical dependencies with the target variables. The result of the feature selection process is to reduce the number N_f of candidate features to a much smaller number $N_r << N_f$ of most relevant features.

The feature selection process yields a set of dimensionally reduced vectors, $\mathbf{x}_i \in \mathrm{R}^{N_r}$, i=1,…M_t. We refer to the set $D = \{(\boldsymbol{x}_i, y_i),\ i = 1,2,...,M_t\}$ as the *training set*. Each of these reduced vectors correspond to a point in an N_r - dimensional feature space. Ideally, these points should cluster into distinct non-overlapping regions in the feature space, corresponding to the respective age groups. In practice however, the clusters may overlap somewhat, so that feature vectors from a few subjects of one age group will map into the cluster of another group, resulting in a classification error corresponding to those subjects. The selection of "better" features; i.e., features with greater statistical dependence on the outcome variable, leads to the formation of tighter clusters with smaller variances and with greater separation between the means of the clusters of different classes, resulting in improved performance.

The reduced feature vectors are fed into a classifier for classification. Generally speaking, the classification process may be viewed as a mapping $f(\mathbf{x})$: $\mathrm{R}^{N_r} \rightarrow y \in C$, between the input feature vector **x** of a test subject and the subject's corresponding age group. Given a set of training patterns where the subject age groups are known, the objective in implementing the classifier is to determine the function *f*. There are many methods of determining the function *f*, which result in different classifiers, e.g., (Vapnik,

1998; Haykin, 2008; Theodoridis and Koutroumbas, 2008). A summary of some classification methods that performed well in the present application is described in subsection 2.5.

2.3 Computing candidate features

For this study, the set of candidate features consists of the DWT of the first- and second-order cumulant functions extracted from the ERP data. Cumulants are average (statistical) quantities and therefore have less inter-trial variance than the ERP signal itself. First order cumulants correspond to the (time-varying) mean value of the signals averaged over all trials and over all electrodes in each region of the scalp, as described above. Second-order cumulants consist of the cross-correlation functions of the averaged signals between respective regions. These cumulants are defined as follows:

1) First order cumulant: $C_{\mathrm{X}}^{1}(n) = m_{\mathrm{x}}(n),\ n = 1,2,....N_e$ (averaged signal of all the sensors in region X)

2) Second order cumulant: $C_{\mathrm{XY}}^{2}(k) = \sum_{n} m_{\mathrm{X}}(n) m_{\mathrm{Y}}(n+k),\ |k| = 1,2,...,N_e - 1$

where $n = 1,\ldots,N_e$ and $m_{\mathrm{X}}(n)$ is the time-varying signal obtained by averaging over all trials and all electrodes in region X**.** The quantities X and Y represent different regions on the scalp: X and Y $\in$ {"FR", "FL" (frontal right and left), "CR", "CL"(Central right and left), "PR", "PL" (Parietal right and left), "OR", "OL" (Occipital right and left), "TR", "TL" (Temporal right and left)}. Since the signal in each region is 1.9 sec long (corresponding to $N_e = 475$ samples) from 0 to 1.9 sec, the duration of each second-order cumulant function is 3.8 sec; i.e., from -1.9 sec to 1.9 sec.

The cumulant sequences themselves are not very efficient as reduced features. However, their wavelet coefficients are much more discriminative as features for this study. The DWT is well known to be effective for compression of signals. Since compression and feature selection are very closely connected entities, it is natural to consider the use of wavelet coefficients as features. The wavelet decomposition is relevant for non-stationary signals and may be interpreted as the time variation of a frequency decomposition of the signal.

The wavelet decomposition and the coherence function corresponding to a second-order cumulant sequence are both frequency domain representations of the EEG signal. The power contained within a

wavelet sequence at a particular frequency band is within a constant multiple of the power contained in the coherence function, over the same band. Since the spectral coherence function between two brain regions at a specific frequency is indicative of synchronization between these regions at that frequency, the power level of the wavelet sequence is also indicative of the same synchronization.

Selection of the appropriate wavelet and the number of decomposition levels is very important in the analysis of signals using the DWT. In this study, a 5-level wavelet decomposition, corresponding to detail components $d1-d5$ and one final approximation component $a5$ (Vetterli and Kovacevic, 1995), was found to yield satisfactory performance. Since the EEG signals are filtered within the band 0.5 – 20 Hz, whereas the Nyquist frequency is at 125 Hz, there are no frequency components of interest in the band 20 – 125 Hz. Therefore, only the detail components ($d3$-$d5$) and the approximation wavelet coefficients ($a5$), which represent the band 0.5 – 20 Hz, are used in subsequent analyses.

The smoothing property inherent in the Daubechies wavelet of order 2 (db2) made it most suitable for use in our application. In our experiments, the total number of candidate features, which are the wavelet coefficients corresponding to the various cumulant sequences, is $N_f = 6330$.

2.4 Feature selection

We use a feature selection procedure based on mutual information (Cover and Thomas, 1991). A useful procedure is to select features that are both relevant (i.e., have high mutual information with the target variables) but also have minimum mutual redundancy. In this respect, we use the suboptimal greedy algorithm of Peng et al. (2005). Suppose that the set of N_r best selected features is denoted by $\mathbf{A}$, and the set of all N_f available features is denoted by $\tilde{\mathbf{X}}$. The first member of $\mathbf{A}$ is the feature with maximum mutual information with the target value y. Then, suppose we already have $\mathbf{A}_{m-1}$, the feature set with m-1 best features. The task is to select the mth feature from the remaining set $\tilde{\mathbf{A}} = \left\{\tilde{\mathbf{X}} - \mathbf{A}_{m-1}\right\}$. This can be done by solving the following optimization problem which implements a trade-off between maximum relevance and minimum redundancy (MRmR)

$$\mathbf{x}_m = \arg\max_{\mathbf{x}_j \in \tilde{\mathbf{A}}} \mu(\mathbf{x}_j) = \arg\max_{\mathbf{x}_j \in \tilde{\mathbf{A}}} \left\{ M(\mathbf{x}_j, y) - \frac{\eta}{m-1} \sum_{\mathbf{x}_i \in \mathbf{A}_{m-1}} M(\mathbf{x}_j, \mathbf{x}_i) \right\} \quad (1)$$

where $\eta > 0$ is a regularization or trade-off parameter and $M(a,b)$ is the mutual information between the random variables *a* nd *b*. Note that the maximized value $\mu(\mathbf{x}_m)$ with respect to the argument provides an indication of the suitability of the proposed *m*th feature. By evaluating (1) over N_r iterations, we are able to produce a selected set of N_r most relevant features.

In order to improve the performance of the feature selection technique and consequently of the classification methods, these features are normalized to have a maximum absolute magnitude of unity, so that each feature is in the interval [-1, 1]. The selected features are then used to train the classifier to determine the age group of each subject.

In order to avoid choosing features that are dominant in just a few patterns, a leave-one-out (LOO) procedure was used to select the best N_r features. The proposed methodology actually uses two LOO procedures executed in succession. The second is used to evaluate the final performance of the method, as described later in Sect. 2.6. The LOO procedure is an iterative process, where in each iteration, all the data associated with one particular subject is omitted from the training set. The iterations repeat until all subjects have been omitted once. In the proposed feature selection scheme, in each iteration, a list of the best $kN_r, k > 1$ features was determined using the MRmR feature selection procedure. For this study the value of *k* was chosen to be 2. After all iterations are complete, the N_r features with the highest number of repetitions (probability of appearance) among the available lists were selected as the final set of selected features.

The optimal value of the parameter N_r was found by first classifying the three age groups using only the single most relevant feature (i.e., $N_r = 1$) using the MRmR procedure. The entire feature selection procedure described above was then applied repetitively, each time incrementing the value of N_r, until no

further improvement was observed in the resulting classification error. This procedure yielded a value N_r = 18.

2.5 Techniques for classification

In this subsection we give a summary of the classification methods that were found to give good performance in our experiments for predicting the age group of the subjects. These include:

1) The kernelized support vector machine (SVM) as proposed by Vapnik (1995). The kernelization procedure imposes a nonlinear transformation on the feature space in a computationally efficient manner (Cristianini and Shawe-Taylor, 2000). The kernelized version of the SVM was found to result in improved performance for this application. This technique requires specification of a kernel function, which is dependent on the specific data (Vapnik, 1995; Cristianini and Shawe-Taylor, 2000; Cortes and Vapnik, 1995). In this paper, the choice of the kernel function was studied empirically and optimal results were achieved using radial-basis function (RBF) kernel function. The SVM is inherently a binary classifier; however, it can be extended into a multi-class classifier by fusing several of its kind together. In our experiments, we fuse SVM binary decisions using the error correcting output-coding (ECOC) approach, adopted from digital communication theory (Dietterich and Bakiri, 1995; Gluer and Ubeyli, 2007).

2) The fuzzy c-means (FCM) algorithm, which is a method of classification where each point is allowed to belong to two or more classes. This method was developed by Dunn (1973) and improved by Bezdek (1981). This algorithm is an iterative classification method having some advantages with respect to other classifiers, the most prominent of which is its high generalization capacity for a reduced number of training trials.

3) The Multilayer Perceptron neural network (MLPNN) classifier. This is the most commonly used neural-network architecture since it enjoys properties such as the ability to learn and generalize, fast operation, and ease of implementation. One major characteristic of these networks is their ability to find nonlinear surfaces separating the underlying patterns. The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks (Haykin, 1998). We use

the Levenberg–Marquardt algorithm to train the MLPNN. This algorithm combines the best features of the Gauss–Newton technique and the steepest-descent algorithm, but avoids many of their limitations (Hagan and Menhaj, 1994).

2.6 The evaluation procedure

The performance of the proposed methodology was evaluated using a second LOO cross-validation procedure. In each iteration (fold) of the current LOO evaluation procedure, the set of features corresponding to one particular subject is again omitted from the training set. The classifier is trained using the remaining available training set and the structure tested using the omitted subject. The test result is compared to the known result provided by the training set. The process repeats M_t times, each time using a different omitted subject, until all subjects have been omitted/ tested once. The same set of previously-identified features is used in each fold. In this way, considering the small size of our available training set, we can obtain an efficient estimate of the performance of the prediction process. LOO cross validation is useful because it does not waste data and provides an asymptotically unbiased estimate of the averaged classification error probability over all possible training sets (Theodoridis and Koutroumbas, 2008). The main drawback of the leave-one-out method is that it is expensive – the computation must be repeated as many times as there are training set data points.

The classifier design and feature selection procedures require the setting of values for various hyperparameters, such as the regularization constant η in (1) and the kernel parameters. These may be conveniently determined using a nested cross-validation procedure within each fold of the main LOO process, in the manner described by (Varma and Simon, 2006; Guyon and Elisseeff, 2003). A flowchart describing the machine learning process for age discrimination is summarized in Fig. 2.

The classification results provided by the LOO procedure can be used to compute various performance indexes, which are indicative of overall performance. The indexes we have chosen are sensitivity, specificity, and total classification accuracy (TCA). These are defined as follows:

- *Sensitivity:* number of subjects that are truly identified to be in one class divided by the number of subjects that are actually in that class.
- *Specificity:* number of subjects that are truly identified not to be in a particular class divided by the total number of subjects that are actually not in that class.
- *Total classification accuracy (TCA):* number of correct identifications in all classes divided by the total number of subjects.

3. Experimental results

The set of the most relevant features selected by the MRmR procedure is shown in Table 1, sorted in terms of the optimized MRmR value $\mu(\mathbf{x})$ from (1). For example, the first row shows that the most relevant feature is the wavelet coefficient of the averaged first-order cumulant sequence C_{OR}^{1} at the occipital right region in the frequency band FB = 3.90-7.81 Hz (theta band), occurring at time T = 1.19 sec., with an MRmR value of $\mu(\mathbf{x})$ =0.7561. The selection of this feature is an indication that this wavelet coefficient changes significantly with the age, and is thus highly indicative of the subject age group.

A further example is the seventh most relevant feature of Table 1, which is the wavelet coefficient of the second-order cross-correlation cumulant sequence $C_{FL,FR}^{2}$ between the frontal right and frontal left regions in the frequency band FB = 7.81-15.63 Hz (alpha band) occurring at time T = 0.28 sec with an MRmR value of $\mu(\mathbf{x})$ = 0.6642. We have seen that the DWT of a cross-correlation function is closely related to the spectral coherence function between the corresponding regions at a specified frequency band, except that the classical definition of coherence does not provide any variation in time. Coherence between two regions at frequency λ indicates there is neural synchronous activity between these regions at that frequency. Thus, the selection of a DWT coefficient of a cross correlation function as a most relevant feature means that synchronous activity between respective regions at a particular frequency is indicative of age group.

An experiment to demonstrate the statistical stability of the selected features is described next. This is important in order to be confident that the results are not skewed by a small number of infants with

anomalous data. Note that this procedure is distinct from the LOO process used to evaluate performance. The results are shown with respect to the 1st feature in Table 1. Five subjects from one particular age group are chosen at random and the *d*5 wavelet coefficient sequences corresponding to this feature are evaluated for each subject. The sequences from these five subjects are then averaged together. This process is repeated 40 times for each of the three age groups, where each time a different set of five subjects is randomly chosen. The resulting averaged sequences are shown in Figs. 3(a)-(c) for the 6-, 12-month-olds and adult age groups, respectively. Fig. 3(d) shows the averaged wavelet coefficients over all subjects in each group. Note that the 1st feature is the value of these sequences at T = 1.19 sec (recall that the four stimulus tones occur every 300 msec with the first tone starting at t = 200 ms in Fig. 3). It may be seen from this figure that the standard deviations of the traces (at T=1.19 sec) are small in comparison to the differences between the traces of the respective age groups, even considering the averaging over the five subjects. Thus, we conclude that this feature is sufficiently statistically stable and provides significant discrimination between the age groups for the particular ERP stimulus used in this experiment. From Fig. 3, it is evident that for the 6-month age group, this feature has a small negative value, a large negative value for the 12-month group, and a large positive value for the adult group. It must be noted that the joint discriminating capability of the combined $N_r = 18$ selected features is significantly improved over the case where only one feature is used; i.e., the statistical behaviour of only this one feature is not an indication of the overall performance of the proposed methodology. Corresponding plots from other brain regions also show similar statistically stable behaviour, and therefore other features likewise provide significant discrimination capabilities.

The overall joint information hidden in the collective of all 18 of these selected features renders the best prediction performance. However, for illustrative purposes only, Fig. 4(a) shows the clustering behaviour of the feature vectors from the respective age groups. This figure was generated by projecting the 18-dimensional feature space onto the first two major principal components for 58 subjects using the principal component analysis (PCA) method. As the figure shows, the three age groups are clearly separated. This supports the assertion that the ERP can be used to determine the age group of the subjects.

Note that even though excellent performance is demonstrated with this 2-dimensional representation, better overall performance is obtained in the $N_r = 18$ dimensional feature space.

A further example showing the behaviour of the selected features is shown in Fig. 4(b). This figure shows the average value of the features between all the subjects in each age group. It may be noted that for most of the selected features, the values of the features for adults and 12-month-olds tend to be large and of opposite polarity, while the corresponding feature for 6-month-olds tends to be small in magnitude.

The classification performance of the proposed methodology for age determination is shown for various classifier structures in Table 2. The MLPNN classifier is used for comparison purposes since it is a very well-known form of classifier (Haykin, 1998). In the hidden layer of the MLPNN, 30 neurons were used. According to Table 2, the SVM and FCM methods perform well in this application, with classification performances above 94%. This verifies the hypothesis that the ERP can yield features that discriminate age group with high reliability.

4. Discussion and conclusions

4.1 Summary

This study proposes a method to determine the age category of 6-month-olds, 12-month-olds, and adults from their ERP responses to a 4-note melody based on modern machine learning principles. Training data from the ERP signals of the three age groups are used to build a classifier, which determines the age group of the subject. The process consists of the following components: feature extraction by computing the wavelet coefficients of the first and second order cumulant sequences, a feature selection procedure where the most statistically relevant features are selected from the set of extracted features, and a classification procedure using classifiers trained on the reduced features.

The feature reduction process uses a “mutual information criterion”, in which the most relevant discriminating features are selected among all the available features, with the condition that they should also satisfy a minimum redundancy criterion. Three different types of classifiers were evaluated. The multiclass SVM and fuzzy C-mean classifiers show more than 94% performance while the performance of

MLPNN was not as high. In addition, we used a low dimensional representation of the feature space using the PCA method that provides a useful tool for visualization of the classification process.

The proposed method of feature selection is in contrast to previous approaches for categorizing subjects according to their ERP components. These methods hypothesize beforehand that a single feature may be discriminative, and then verify or reject this hypothesis by experiment. In contrast, our proposed feature selection method finds a small number of maximally discriminative features that are *automatically* identified from a very large list of candidate features. Thus our method can potentially identify salient features that could be missed using previous methods. ~~and chooses only those that are most discriminative.~~

It should be noted that the top 18 features described in Table 1 are not unique. Due to the rich redundancy of the candidate features, other selected feature sets could be chosen with almost equal MRmR values. An interesting topic for further investigation is to explicitly include various parameters relating to the ERP components (such as component intensity, latency, duration, etc.) in the list of candidate features, to determine whether they are chosen as selected features.

4.2 Over-training

Over-training is always an issue in any machine learning application. Over-training happens when the feature selection and classifier design processes over-adapt to the specific training set, with the result that the resulting structure performs well with the given training set, but does not generalize well to new samples. We now present examples that suggest over-training is not a dominant phenomenon in this study. First, the behaviour shown in Fig. 4(a) shows clean separation of the clusters representing each class, which means that good classification performance can be obtained with boundaries in the form of low-dimensional hyperplanes. This suggests the boundaries have not over-adapted to the specific training set, and therefore the classifier structure should behave well with new data. The second demonstration is based on the argument that when the dimension of the feature space is comparable to the number of training samples, over-training may exist. In the first two columns of Table 3, we show performance results corresponding to those shown in the last column of Table 2, except that we use different values of N_r. It is seen that performance is not overly sensitive to this parameter. Particularly, performance is not seriously

degraded when N_r is reduced to 12, which is approximately 1/5 of the total number of training samples. Thus, the proposed structure behaves well when the dimension of the feature space is significantly lower than the number of training samples, further suggesting that over-training is not a dominant consideration in this study.

An additional demonstration involves testing variations of the same training set. In this procedure, we used 80% of the subjects in each age group for training and the remaining 20% of the subjects for testing. A hundred experiments with different randomly selected training and test subjects were carried out and the average performance is reported in the third column of Table 3. As the table shows, the performance of the classifiers do not change significantly in comparison to that shown in Table 2, suggesting over-training has not occurred. The final point with regard to over-training concerns feature selection. The regularized feature selection method described in subsection 2.4 is specifically chosen to avoid the situation where a few training samples dominate the feature selection process.

4.3 Neurophysiological interpretation of the selected features

The optimality of our proposed feature selection procedure suggests that these selected features are highly indicative of the underlying neurophysiological processes that accompany development. A complete understanding of the clues these features provide with respect to neural development is beyond the scope of this paper and remains a topic for future work. Nevertheless, we present some examples and observations in the following paragraphs that provide some limited insight in this respect.

Features 1-6, 8, 9, 11, 14, 15 of Table 1 are all wavelet coefficients extracted from first-order cumulant sequences of the ERP waveforms in the theta band (3.9-7.8 Hz), and therefore probably capture age differences in traditional ERP components such P1, N1 and P3 that fall within this frequency range. Most of the first order features (features 1, 2, 3, 8, 9, 11) occur at time T = 1.19 sec at widespread regions (OR, OL, TL, FL, CL, and TR) across the brain. The left-hand side (panels (a)-(d)) of Fig. 5 shows the first-order cumulant sequences from the FL, CL, TL and OL regions. These sequences are equivalent to the traditional ERP waveforms. The right-hand side (panels (e)-(h)) of the figure shows the corresponding wavelet sequences in the 3.90-7.81 Hz frequency band. These specific sequences were chosen because

they contain many of the selected features. The maximal distinction between the age groups is clearly evident from these wavelet sequences at time T = 1.19 sec. T = 1.19 sec is about 100 ms after onset of the final (fourth) tone of the melodies. Because the fourth tone of the melody was occasionally played incorrectly, even though we did not analyze incorrectly played trials, attention was likely directed to this time period. Examination of the averaged waveforms (i.e., the first-order cumulant sequences) from Fig. 5(a)-(d) reveals that adults show an N1/P2 complex after the fourth tone, with the N1 centred around 100 ms after tone onset, as shown in Fig. 5(a). Because the N1/P2 complex is largely within the 3.9-7.8 Hz frequency range, the wavelet sequences for adults show significant energy in this band. The 12-month-olds also show significant energy in this band with a reversed polarity at T = 1.19 sec relative to adults. The N1/P2 component does not appear in the cumulant sequences for the 6-month olds, which accounts for the diminished energy of the wavelets in the 3.9-7.8 Hz band for this age group. Figure 5 also shows that the wavelet feature patterns and the original ERPs across all ages reverse in polarity at the frontal and central regions compared to occipital and temporal at T = 1.19 sec, consistent with dipolar generators of this electrical activity in the auditory cortices (Trainor, 2008).

First order features also occur at time T = 0.71 sec in the frequency band FB = 3.9-7.8 Hz at the OR (feature 4, Fig. 3(d)), FL (feature 5, Fig. 5(e)), CL (feature 6, Fig. 5(f)), and OL regions (feature 15, Fig. 5(h)). Note that the three age groups show very different wavelet coefficients at these regions at this time. T=0.71 sec is about 200 ms after the onset of the 2nd tone. Here the corresponding adult ERP waveforms shown in Fig. 5 consistently show a frontal and central positivity whereas 12-month-olds show a negativity at these times (Figs. 5(e) and (f)). Six-month-olds have very little energy in this band, resulting in low-level wavelet coefficients. These features also reverse polarity from the front to the back of the head, (see Figs. 5(e) vs. (h)) again consistent with generators of activity in the auditory cortex.

The wavelet coefficients extracted from second order cumulant sequences are all in the alpha band (7.81-15.63 Hz). As previously discussed, this suggests that alpha-band synchronization between regions is an additional neural condition that changes with development. For example, Figs. 6(a) and (c) show the cross-correlation and corresponding wavelet sequences between the frontal left and right regions (feature

7), whereas Figs. 6(b) and (d) show similar plots between the temporal right and occipital right regions (feature 12). The cross-correlation sequences for all three age groups show the largest peak near zero. For feature 7, this means that the two hemispheres are quite closely in synch with no time delay. For all three age groups, the wavelet sequences within this band exhibit narrow-band oscillatory behaviour, with a centre frequency that varies with age. Thus the wavelet coefficients for the three age groups are in-phase at some delays and out-of-phase at others, allowing there to exist a delay value at which the wavelet sequences of the three age groups are maximally different, and therefore qualify as a selected feature in Table 1. The change in the frequency of the oscillatory characteristic of these wavelet sequences with age is an indication that changes in synchrony between the left and right frontal regions (feature 7) and temporal right and occipital right regions (feature 12) are an indication of developmental maturation of the human brain.

Although we cannot know for sure what neurological developments are associated with the age differences that are apparent, the first order cumulants are likely associated with short-range maturation of connections between neurons. It is known from autopsy studies of human brain tissue that myelination and neurofilament expression increase in auditory areas during infancy, which enables faster and more efficient connections between neurons with increasing age (Huttenlocher and Dabholkar, 1997; Moore and Guan, 2001). The differences in synchrony between brain regions uncovered in the second order cumulants are perhaps more interesting in that there are few previous studies showing developmental EEG differences related to changes in long-range connections, but this development is crucial for optimal brain functioning (e.g., Casanova et al., 2009; Keary et al., 2009; Thatcher et al., 2008).

4.4 Conclusions

In sum, we have shown that the present approach of using a machine learning procedure that does not require prior hypotheses for uncovering features that distinguish maturational age has the potential to uncover new theoretical understanding of maturation changes in long-range synchrony. It also opens the possibility of devising a clinical test that can compare the chronological and maturational ages of individual subjects in order to determine whether an infant is developing normally or experiencing

significant delay. In the present study, we compared only three ages. It remains for further study to determine how fine-grained the classification by age can be made.

References

Bezdek JC. Pattern recognition with fuzzy objective function algorithms, New York: Plenum Press, 1981.

Casanova MF, El-Baz A, Mott M, Mannheim G, Hassan H, Fahmi R, et al. Reduced gyral window and corpus callosum size in autism: possible macroscopic correlates of a minicolumnopathy. J Autism Dev Disord 2009; 39:751-764.

Čeponiené R, Kushnerenko E, Fellman V, Renlund M, Suominen K, Näätänen R. Event-related potential features indexing central auditory discrimination by newborns. Cogn Brain Res 2002; 13: 101-113.

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(2): 320-38.

Cortes C, Vapnik VN. Support vector networks. Mach Learn 1995; 20(3): 273-297.

Cover TM, Thomas JA. Elements of information theory, New York: Wiley, 1991.

Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods, 1st ed. Cambridge: Cambridge University Press, 2000.

Curtin S, Werker JF. The perceptual foundations of phonological development. In: Gaskell MG, editor. The Oxford Handbook of Psycholinguistics. Oxford: Oxford University Press, 2007: 579-599.

Davidson RJ. Anterior cerebral asymmetry and the nature of emotion. Brain Cogn 1992; 20(1): 125-151.

De Boer T, Scott LS, Nelson CA. Methods for acquiring and analyzing infant event-related potentials. In: De Haan M, editor. Infant EEG and event-related potentials: Studies in developmental psychology. New York: Psychology Press, 2007: 5-37.

De Haan M, editor. Infant EEG and event-related potentials: Studies in developmental psychology, New York: Psychology Press, 2007.

Dietterich TG, Bakiri G. Solving multiclass learning problems via error-correcting output codes. J Artif Intell Res 1995; 2: 263-286.

Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cyber 1973; 3(3): 32-57.

Friedrich H, Friederici AD. Early N400 development and later language acquisition. Psychophysiol 2006; 43(1): 1-12.

Ghosh-Dastidar S, Adeli H, Dadmehr N. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. IEEE Trans Biomed Eng 2008; 55(2): 512-518.

Gluer I, Ubeyli ED. Multiclass support vector machines for EEG-Signals classification. IEEE Trans Inf Technol Biomed 2007; 11(2): 117-126.

Greene BR, De Chazal P, Boylan GB, Connolly S, Reilly RB. Electrocardiogram based neonatal seizure detection. IEEE Trans Biomed Eng 2007; 54(4): 673-682.

Guttorm TK, Leppänen PHT, Richardson U, Lyytinen H. Event-Related Potentials and Consonant Differentiation in Newborns with Familial Risk for Dyslexia. J Learn Disabil 2001; 34(6): 534-544

Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003; 3: 1157-1182.

Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Net 1994; 5(6): 989-993.

Hannon EE, Trainor LJ. Music acquisition: effects of enculturation and formal training on development. Trends Cogn Sci 2007; 11: 466- 472.

Haykin S. Neural networks: A comprehensive foundation, 2nd ed. Prentice Hall, 1998.

Haykin S. Neural networks and learning machines, 3rd ed. Prentice Hall, 2008.

He C, Hotson L, Trainor LJ. Mismatch responses to pitch changes in early infancy. J Cogn Neurosci 2007; 19(5): 878-892.

He C, Hotson L, Trainor LJ. Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. Eur J Neurosci 2009a; 29: 861-867.

He C, Hotson L, Trainor LJ. Maturation of cortical mismatch mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. Neuropsychol 2009b; 47: 218-229.

He C, Trainor LJ. Finding the pitch of the missing fundamental in infants. J Neurosci 2009; 29: 7718-7722.

Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. J Comp Neurol 1997; 387(2): 167-178.

Keary CJ, Minshew NJ, Bansal R, Goradia D, Fedorov S, Keshavan MS, et al. Corpus callosum volume and neurocognition in autism. J Autism Dev Disord 2009; 39: 834-41.

Khodayari-Rostamabad A, Hasey GM, MacCrimmon DJ, Reilly JP, de Bruin H. A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. Clin Neurophysiol 2010; 121(12): 1998-2006.

Knott VJ, La Belle A, Jones B, Mahoney C. EEG coherence following acute and chronic clozapine in treatment-resistant schizophrenics. Exp Clin Psychopharmacol 2002; 10(4): 435-444.

Krajča1 V, Petránek1 S, Mohylová J, Paul K, Gerla V, Lhotská L. Neonatal EEG sleep stages modeling by temporal profiles. Comp Aided Syst Theory, Eurocast 2007; 195-201.

Kropotov JD, Näätäen R, Sevostianov AV, Alho K, Reinikainen K., Kropotova OV. Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex. Psychophysiol 1995; 32(4): 418-422.

Kuhl PK. Linking infant speech perception to language acquisition: Phonetic learning predicts language growth. In McCardle P, Colombo J, Freund L, editors. Infant pathways to language: Methods, models, and research directions. Erlbaum: New York, 2008: 213-243.

Kushnerenko E, Ceponiene R, Balan P, Fellman V, Naatanen R. Maturation of the auditory change-detection response in infants: a longitudinal ERP study. NeuroReport 2002a; 13(15): 1843-1848.

Kushnerenko E, Ceponiene R, Balan P, Fellman V, Huotilainen M, Naatanen R. Maturation of the auditory event-related potentials during the first year of life. NeuroReport 2002b; 13(1): 47-51.

Kwon JS, Youn T, Jung HY. Right hemisphere abnormalities in major depression: quantitative electroencephalographic findings before and after treatment. J Affect Disord 1996; 40(3): 169-173.

Moore JK, Guan YL. Cytoarchitectural and axonal maturation in human auditory cortex. J Assoc Res Otolaryngol 2001; 2: 297-311.

Morr ML, Shafer VL, Kreuzer JA, Kurtzberg D. Maturation of Mismatch Negativity in Typically Developing Infants and Preschool Children. Ear Hear 2002; 23: 118-136.

Mourad N, Reilly JP, De Bruin H, Hasey G, MacCrimmon D. A simple and fast algorithm for automatic suppression of high-amplitude artifacts in EEG data. ICASSP 2007; 1: I-393-I-396.

Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin Neurophysiol 2007; 118: 2544-2590.

Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005; 27(8): 1226-1238.

Picton TW, Alain C, Otten L, Ritter W, Achim A. Mismatch negativity: Different water in the same river. Audiol Neurootol 2000; 5(3-4): 111-139.

Taylor MJ, Baldeweg M. Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. Dev Sci 2002; 5(3): 318-334.

Tew S, Fujioka T, He C, Trainor LJ. Neural representation of transposed melody in infants at 6 months of age. Ann N Y Acad Sci 2009; 1169: 287-290.

Thatcher RW, North DM, Biver CJ. Development of cortical connections as measured by EEG coherence and phase delays. Hum Brain Mapp 2008; 29(12): 1400-1415.

Theodoridis S, Koutroumbas K. Pattern recognition, 4th ed. Academic Press, 2008.

Trainor LJ, Samuel SS, Galay L, Hevenor SJ, Desjardins RN, Sonnadara R. Measuring temporal resolution in infants using mismatch negativity. NeuroReport 2001; 12: 2443-2448.

Trainor LJ, McFadden M, Hodgson L, Darragh L, Barlow J, Matsos L, et al. Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. Int J Psychophysiol 2003; 51:5-15.

Trainor LJ. Event-related potential (ERP) measures in auditory developmental research, In: Schmidt LA, Segalowitz SJ, editors. Developmental psychophysiology: Theory, systems and methods. New York: Cambridge University Press, 2008: 69-102.

Vapnik VN. The nature of statistical learning theory, NewYork: Springer-Verlag, 1995.

Vapnik VN. Statistical learning theory, New York: Wiley, 1998.

Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006; 7(1): 91.

Vetterli M, Kovacevic J. Wavelets and subband coding, Prentice Hall, 1995.

List of figure captions:

Fig. 1. Electrode groupings in the HydroCel GSN net. Ninety out of 128 electrodes were selected to be divided into ten regions (frontal right and left (FR and FL), central right and left (CR and CL), parietal right and left (PR and PL), occipital right and left (OR and OL), and temporal right and left (TR and TL)). Each region included 8 to 10 channels.

Fig. 2. Flow chart of the proposed age discrimination procedure.

Fig. 3. Averaged wavelet sequences over five randomly-chosen subjects of the first selected feature of Table 1, for the cases of (a) 6-month-old infants, (b) 12-month-old infants and (c) adults. The process was repeated over 40 random trials. (d) The averaged wavelet coefficients over all subjects in each group. The first selected feature is the value of this sequence at T=1.19 sec where it can be seen that the ages groups are maximally different.

Fig. 4. (a) Subject-wise scatter plot of the feature space projected onto the first two major principal components, (b) the mean values of the features between all the subjects in each group.

Fig. 5. The averaged ERP signal (1^{st}-order cumulants) (left) and the corresponding wavelet coefficients (right) over all subjects in each group for the (a), (e) frontal left, (b),(f) central left, (c),(g) temporal left and (d),(h) occipital left regions, respectively.

Fig.6. The 2^{nd}-order cumulant sequences (left) and the corresponding wavelet sequences (right) between (a), (c) frontal left and right and (b), (d) temporal right and occipital right regions, respectively in frequency band of FB = 7.81-15.63 Hz.

Table 1

List of the $N_r = 18$ selected features used to predict the age group of subjects and their MRmR criteria value $\mu(\mathbf{x})$, where "FB" and "T" denote the frequency band and the time for each wavelet coefficient, respectively.

Feature #	Feature	MRmR
1	C^1_{OR}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.7561
2	C^1_{OL}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.7526
3	C^1_{TL}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.7422
4	C^1_{OR}, FB = 3.90-7.81 Hz, T = 0.71 sec	0.7319
5	C^1_{FL}, FB = 3.90-7.81 Hz, T = 0.71 sec	0.7008
6	C^1_{CL}, FB = 3.90-7.81 Hz, T = 0.71 sec	0.6694
7	$C^2_{FL,FR}$, FB = 7.81-15.63 Hz, T = 0.28 sec	0.6642
8	C^1_{FL}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.6626
9	C^1_{CL}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.6617
10	C^1_{FL}, FB = 7.81-15.63 Hz, T = 0.61 sec	0.6467
11	C^1_{TR}, FB = 3.90-7.81 Hz, T = 1.19 sec	0.6453
12	$C^2_{TR,OR}$, FB = 7.81-15.63 Hz, T = -0.22 sec	0.6357
13	$C^2_{FR,OL}$, FB = 7.81-15.63 Hz, T = -0.59 sec	0.6217
14	C^1_{TR}, FB = 3.90-7.81 Hz, T = 0.36 sec	0.6185
15	C^1_{OL}, FB = 3.90-7.81 Hz, T = 0.71 sec	0.6179
16	$C^2_{CL,CR}$, FB = 7.81-15.63 Hz, T = -0.22 sec	0.6095
17	C^1_{OL}, FB = 7.81-15.63 Hz, T = 1.10 sec	0.6086
18	$C^2_{OL,OR}$, FB = 7.81-15.63 Hz, T = -0.09 sec	0.6081

Table 2
Comparison of the performance among different classifiers for predicting the age of subjects using all selected features, for $N_r = 18$.

Method	Classes	6-month	12-month	Adults	Sensitivity	Specificity	TCA
MLPNN	6-month	26	2	1	89.7%	82.8%	84.5%
	12-month	4	15	0	78.9%	92.3%	
	Adults	1	1	8	80%	97.9%	
SVM	6-month	28	1	0	96.6%	93.1%	94.8%
	12-month	2	17	0	89.5%	97.4%	
	Adults	0	0	10	100%	100%	
FCM	6-month	27	1	1	93.1%	96.5%	94.8%
	12-month	1	18	0	94.7%	97.4%	
	Adults	0	0	10	100%	95.9%	

Table 3
Comparison of performance among different classifiers in predicting the age of subjects under varying conditions. The first two columns show the performance obtained from the LOO cross-validation procedure, for different values of N_r. The third column shows results where all 18 features are used, and 80% of the subjects in each group are used for training, and the remaining 20% are used for evaluation.

Method	TCA using LOO with 12 features	TCA using LOO with 15 features	TCA using 80% of the subjects and all 18 features for training
MLPNN	78.3%	82.1%	81.2%
SVM	88.2%	92.7%	91.5%
FCM	89.7%	92.7%	93.8%

Figure 1

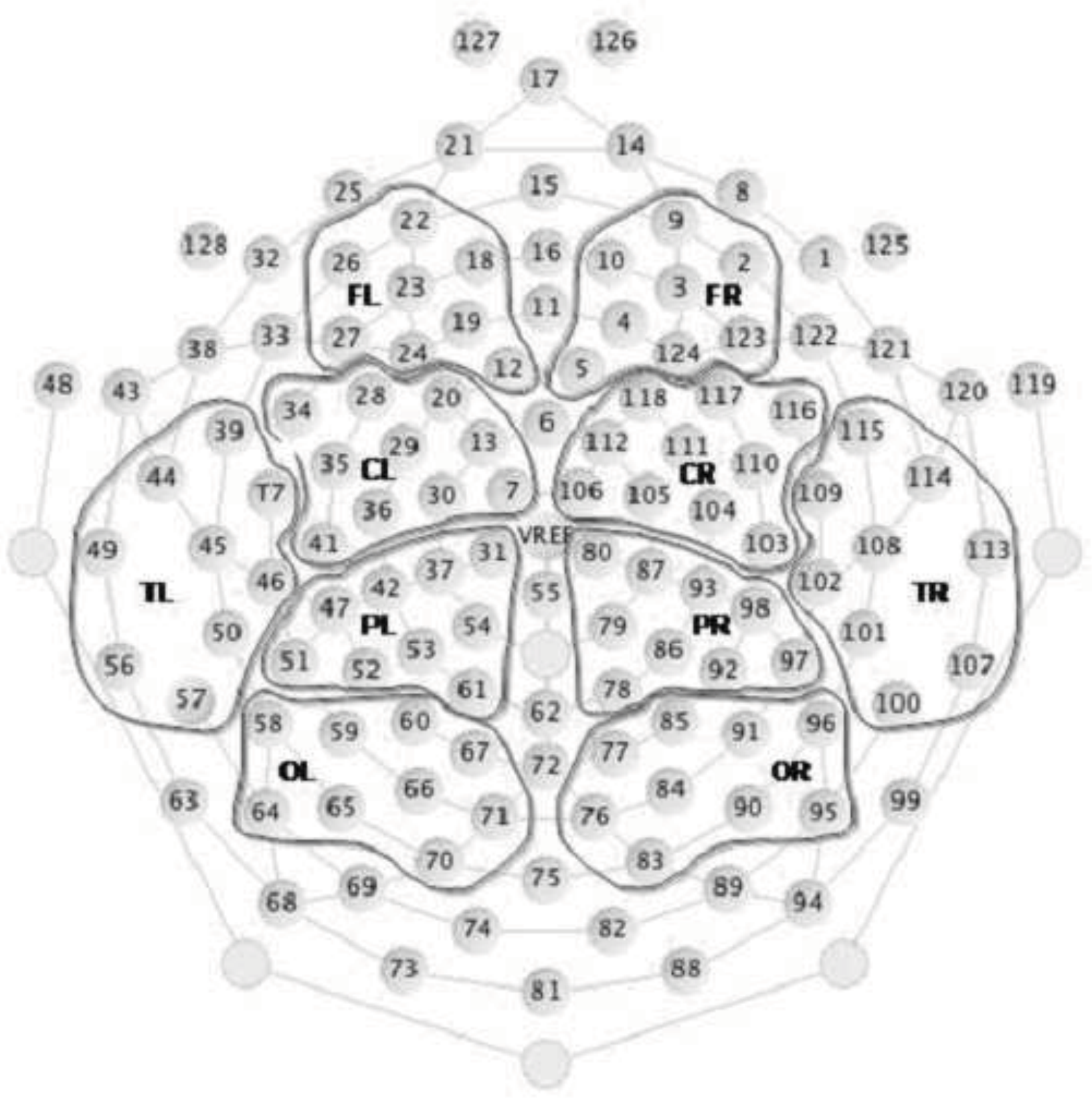

127
126
17
21
14
25
15
8
22
9
128
32
26
18
16
10
2
1
125
FL
23
3
FR
11
4
19
27
123
122
38
33
24
12
5
124
121
48
43
28
20
118
117
116
120
119
34
6
39
29
13
112
111
115
35
CL
CR
110
44
T7
30
7
106
105
109
114
36
104
VREF
41
103
108
113
49
45
31
80
87
37
102
46
42
55
93
TL
TR
47
98
PL
PR
50
54
79
101
53
86
97
56
51
52
92
107
61
78
57
62
100
58
59
60
85
91
96
67
77
72
OL
OR
63
66
84
99
64
65
71
76
90
95
70
83
75
69
89
68
94
74
82
73
88
81

Figure 2

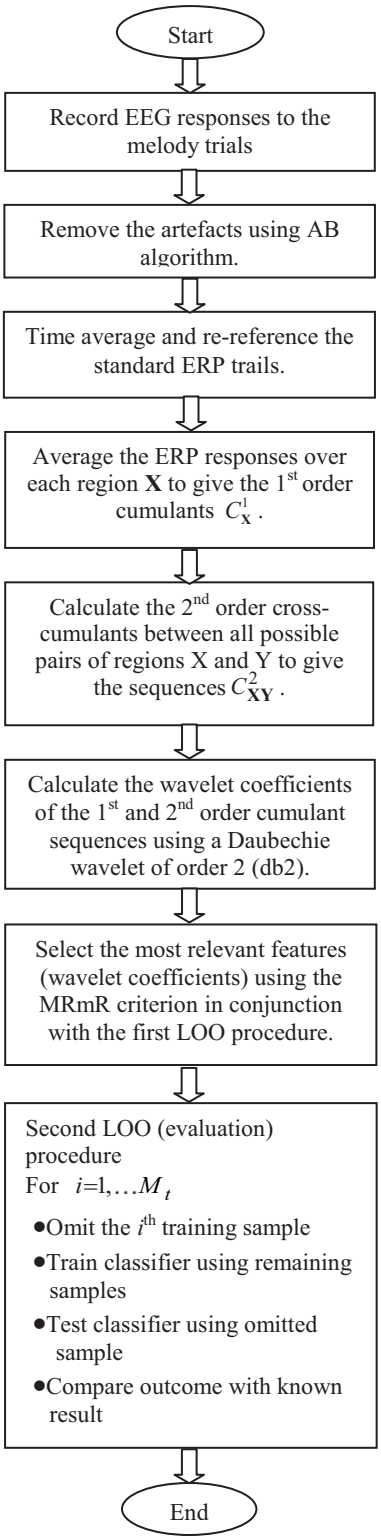

Start
Record EEG responses to the melody trials
Remove the artefacts using AB algorithm.
Time average and re-reference the standard ERP trails.
Average the ERP responses over each region **X** to give the 1st order cumulants $C^1_{\mathbf{X}}$.
Calculate the 2nd order cross-cumulants between all possible pairs of regions X and Y to give the sequences $C^2_{\mathbf{XY}}$.
Calculate the wavelet coefficients of the 1st and 2nd order cumulant sequences using a Daubechie wavelet of order 2 (db2).
Select the most relevant features (wavelet coefficients) using the MRmR criterion in conjunction with the first LOO procedure.
Second LOO (evaluation) procedure
For i=1,…M_t
Omit the i^{th} training sample
Train classifier using remaining samples
Test classifier using omitted sample
Compare outcome with known result
End

Figure 3

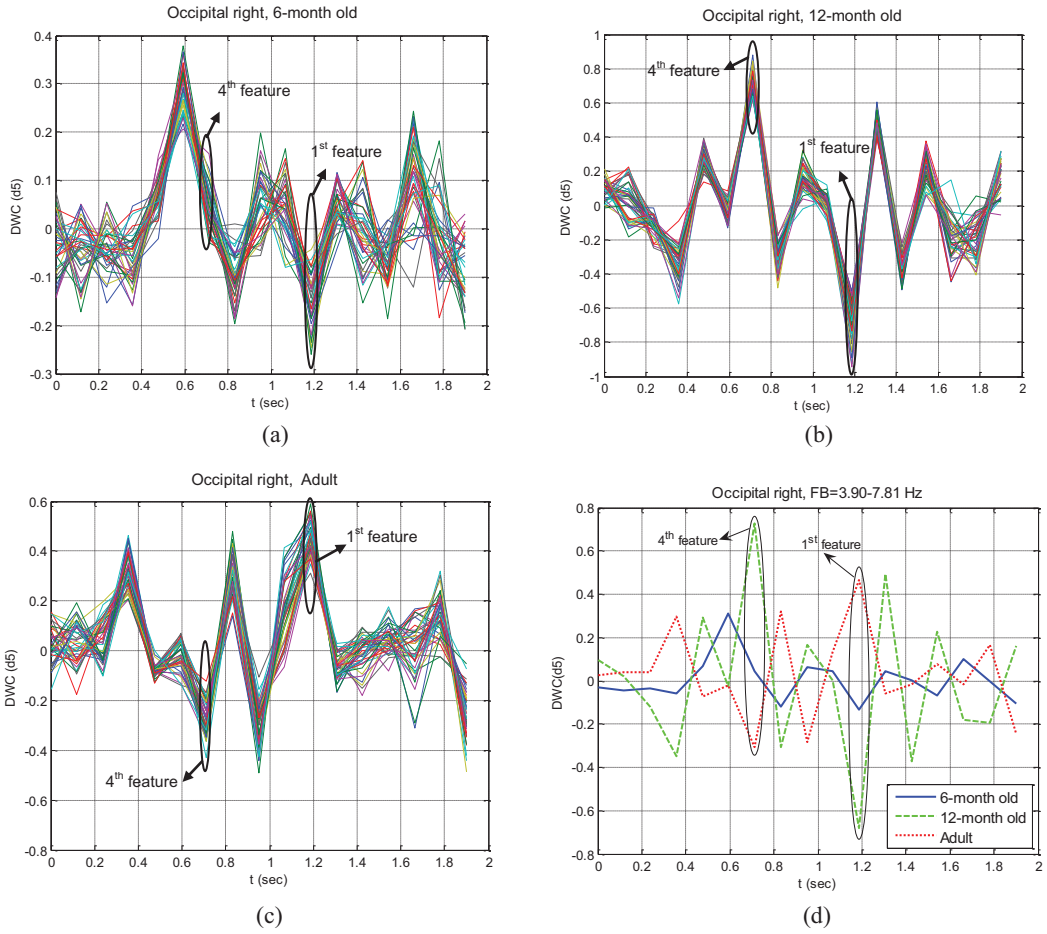
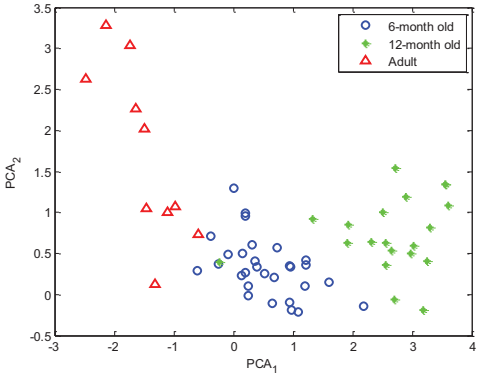
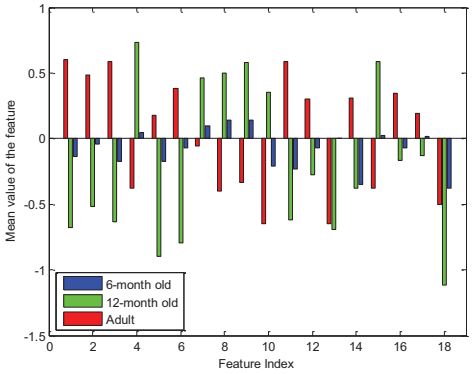

Occipital right, 6-month old
4th feature
1st feature
DWC (d5)
t (sec)
(a)
Occipital right, 12-month old
4th feature
1st feature
DWC (d5)
t (sec)
(b)
Occipital right, Adult
1st feature
4th feature
DWC (d5)
t (sec)
(c)
Occipital right, FB=3.90-7.81 Hz
4th feature
1st feature
DWC(d5)
6-month old
12-month old
Adult
t (sec)
(d)

Figure 4


6-month old
12-month old
Adult
PCA2
PCA1

(a)


Mean value of the feature
Feature Index
6-month old
12-month old
Adult

(b)

Figure 5

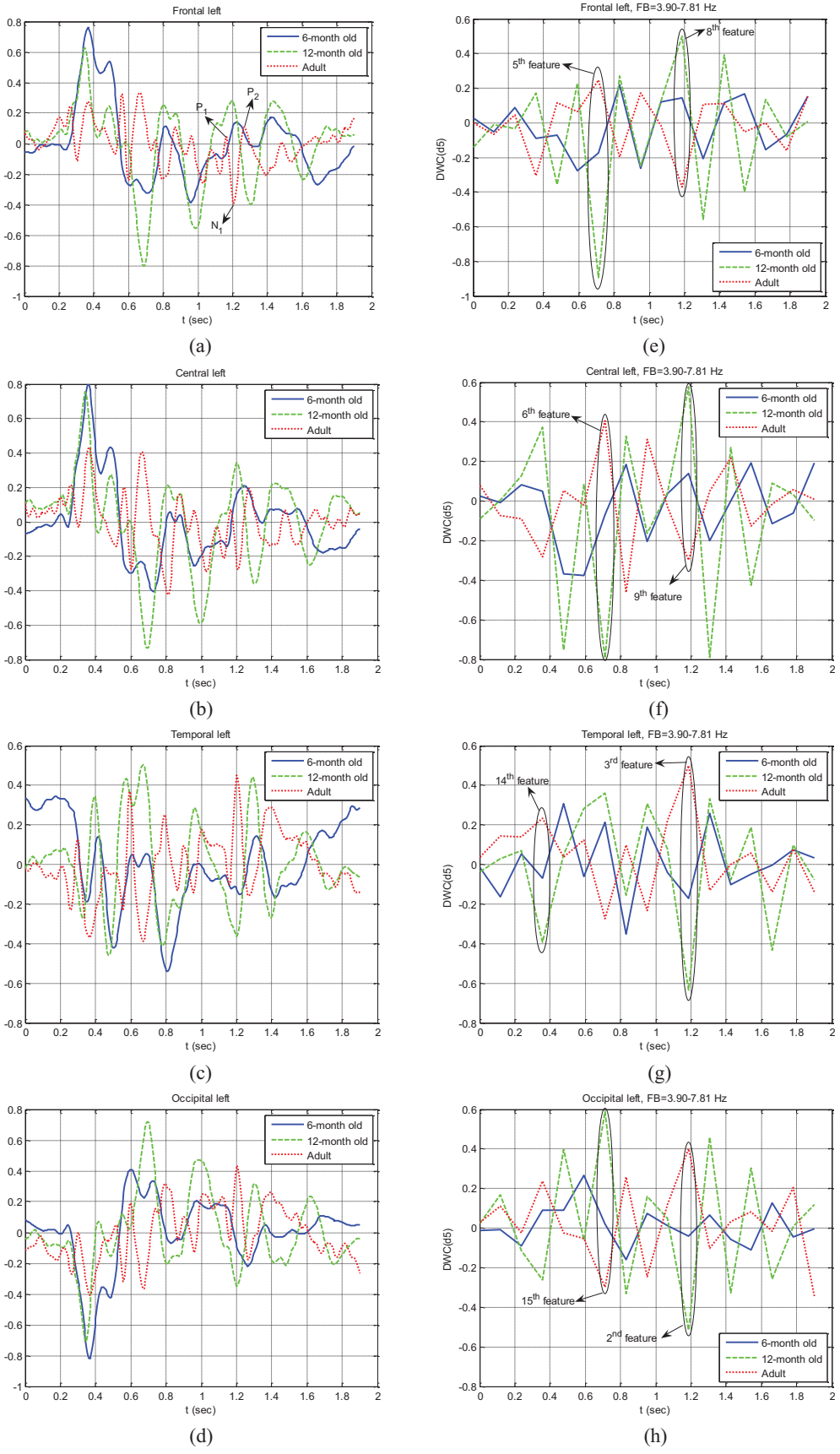
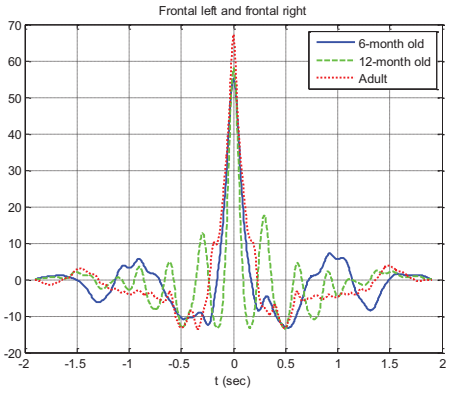
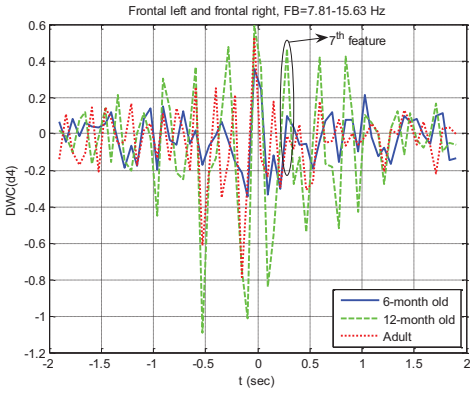

Frontal left
6-month old
12-month old
Adult
P1
P2
N1
t (sec)
(a)
Central left
6-month old
12-month old
Adult
t (sec)
(b)
Temporal left
6-month old
12-month old
Adult
t (sec)
(c)
Occipital left
6-month old
12-month old
Adult
t (sec)
(d)
Frontal left, FB=3.90-7.81 Hz
8th feature
5th feature
DWC(d5)
6-month old
12-month old
Adult
t (sec)
(e)
Central left, FB=3.90-7.81 Hz
6th feature
9th feature
DWC(d5)
6-month old
12-month old
Adult
t (sec)
(f)
Temporal left, FB=3.90-7.81 Hz
3rd feature
14th feature
DWC(d5)
6-month old
12-month old
Adult
t (sec)
(g)
Occipital left, FB=3.90-7.81 Hz
15th feature
2nd feature
DWC(d5)
6-month old
12-month old
Adult
t (sec)
(h)

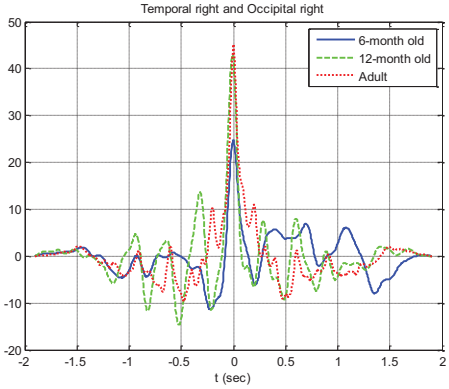
Figure 6



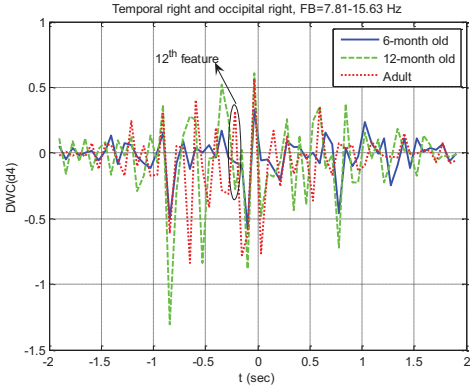
(a)



(c)



(b)



(d)