Published in final edited form as:

Clin Neurophysiol. 2013 March ; 124(3): 452–461. doi:10.1016/j.clinph.2012.09.003.

AUTOMATING THE ANALYSIS OF EEG RECORDINGS FROM PREMATURELY-BORN INFANTS: A BAYESIAN APPROACH

Timothy J. Mitchell^{a,*}, Jeffrey J. Neil^{a,b,c}, John M. Zempel^{a,b}, Liu Lin Thio^{a,b}, Terrie E. Inder^{a,b,c}, and G. Larry Bretthorst^c

^aDepartment of Pediatrics, Washington University, St. Louis, MO 63110

^bDepartment of Pediatric Neurology, Washington University, St. Louis, MO 63110

^cDepartment of Radiology, Washington University, St. Louis, MO 63110

Abstract

Objective—To implement an automated analysis of EEG recordings from prematurely-born infants and thus provide objective, reproducible results.

Methods—Bayesian probability theory is employed to compute the posterior probability for developmental features of interest in EEG recordings. Currently, these features include smooth delta waves (0.5-1.5 Hz, >100 μV), delta brushes (delta portion: 0.5-1.5Hz, >100 μV ; “brush” portion: 8-22 Hz, <75 μV), and interburst intervals (<10 μV), though the approach taken can be generalized to identify other EEG features of interest.

Results—When compared with experienced electroencephalographers, the algorithm had a true positive rate between 72% and 79% for the identification of delta waves (smooth or “brush”) and interburst intervals, which is comparable to the inter-rater reliability. When distinguishing between smooth delta waves and delta brushes, the algorithm's true positive rate was between 53% and 88%, which is slightly less than the inter-rater reliability.

Conclusions—Bayesian probability theory can be employed to consistently identify features of EEG recordings from premature infants.

Significance—The identification of features in EEG recordings provides a first step towards the automated analysis of EEG recordings from premature infants.

Keywords

EEG; Bayesian; Neonatal

1. Introduction

Premature infants undergo extensive cerebral maturation in the neonatal intensive care unit (NICU), but they are also vulnerable to brain injury that can result in later motor and cognitive disability (Hack et al., 2002; Wood et al., 2000). Because of the risks to the infant, it is crucial that physicians be able to identify and provide treatment for brain insults as early as possible. When available, electroencephalographic (EEG) recordings play an important role in identifying alterations in cerebral function and they can be obtained continuously at the bedside.

One impediment to the widespread adoption of EEG within the NICU is that it must be interpreted by a skilled electroencephalographer, but access to an electroencephalographer is not available in all NICUs. Additionally, there are numerous subtleties involved in the interpretation of EEG recordings such that even experts may disagree, making EEG analysis somewhat subjective. Furthermore, even when an expert is available, the process of reading an EEG is often time-consuming, resulting in delays of a day or more between recording and diagnosis. Each of these limitations could be addressed by an automated approach to EEG analysis. An algorithm designed to aid in the analysis of EEG would increase access to this vital tool by making EEG practical in NICUs without an electroencephalographer, eliminate subjectivity in the diagnosis, and provide real-time information to physicians. For a further explanation of the value of a quantitative approach, see Niemarkt et al. (2008).

Many attempts to automate the analysis of EEG recordings have dealt with seizure detection (Gotman et al., 1997; Liu et al., 1992; Roessgen et al., 1998; Wilson, 2005; Deburchgraeve et al., 2008; Navakatikyan et al., 2006; Altenburg et al., 2003; Temko et al., 2011; Faust et al., 2010). However, there is a growing body of literature attempting to obtain quantitative measures of the non-seizure EEG. A number of automated techniques have emerged for the classification of background activity either through the detection of the burst-suppression pattern (Löfhede et al., 2010; Palmu et al., 2010a,b) or the detection of interburst intervals (Niemarkt et al., 2010a). Others have used quantitative analysis to identify normative patterns of development. These analyses have used aEEG analysis (Niemarkt et al., 2010b), interburst interval detection (Niemarkt et al., 2010a), frequency analysis (Victor et al., 2005; West et al., 2006; Scher, 1995), amplitude analysis (West et al., 2006), or a combination of many measures including spectral entropy (Korotchikova, 2009). The use of quantitative analyses has led to several interesting findings. Inder et al. (2003) found a correlation between decreased spectral edge frequency and increased severity of white matter injury. Eiselt et al. (1997) compared the spectra between the two hemispheres of the brain and found that the spectral power was higher in the right hemisphere than the left, indicating functional differences between the two sides are present even in the neonate. Korotchikova (2011) found that grades of hypoxic ischaemic encephalopathy could be predicted using a combination of several quantitative EEG measures.

While many automated analyses of EEG recordings have been successful in their attempt to agree with a visual analysis of the EEG provided by electroencephalographers, no automated analyses have been widely adopted to date, meaning more work is needed to bring this valuable tool into the NICU. The primary focus of this study is identifying developmental

landmarks. The landmarks of interest are interburst intervals and delta waves, with delta waves being further subdivided into smooth delta waves and delta brushes. Delta waves are defined as large amplitude ($> 100\mu V$) waves having frequencies between 0.5 and 1.5 Hz. Delta brushes are delta waves that contain superimposed high-frequency “brush” waves, which are typically located near the top of the delta wave and can range anywhere between 8 and 22 Hz in frequency and have amplitudes of 10 to 75 μV . Delta brushes have also been referred to as “beta-delta complexes” or “spindle-like fast waves” (Mizrahi et al., 2004). Smooth delta waves are any delta waves not containing brush activity. Examples of delta waves and delta brushes can be seen in Fig. 1(a), where smooth delta waves are marked “SDW” and the delta brush is marked “DB”. Note that not all smooth delta waves are perfectly smooth, but simply lack brush activity. Delta brushes typically appear around 28 weeks postmenstrual age (PMA) and continue through 39 weeks with a gradual progression from the central regions to the occipital and temporal areas. Their absence from these regions during this period is indicative of abnormal development. However, the presence of delta waves can be a sign of pathology; sustained rhythmic delta activity in the bifrontal or bioccipital regions both represent an abnormal EEG finding (Mizrahi et al., 2004).

Another developmental landmark of interest is the interburst interval, which is a period of low-voltage ($< 10\mu V$) signal during which brain activity is minimal. Figure 1(b) contains an example of an interburst interval, with the arrows marking the beginning and end of the interburst interval. In general, the EEG should become more continuous as PMA increases in the healthy neonate, meaning maturation can be tracked by identifying interburst intervals.

The goal of this project is to provide a first step towards the automation of EEG analysis in premature infants. The algorithm is designed to mimic electroencephalographers by identifying developmental landmarks known to correlate with normal or abnormal neurological development in premature infants within the NICU environment. The landmarks of interest in this project are delta brushes, smooth delta waves, and interburst intervals, though the method of identification can be expanded to identify nearly any wave of interest. The implementation of this algorithm has the potential to offer objective, real-time information to clinicians at the bedside without requiring a specialist to interpret the EEG.

2. Methods

2.1. Bayesian Probability

Bayesian probability theory, a mathematical theory of inference in which probabilities represent degrees of plausibility or a state of knowledge regarding a particular hypothesis, is employed to identify the EEG features of interest. Jaynes (1956, 2003), building on the works of many previous authors such as Cox (1946) and Jeffreys (1939), worked to fully develop the theory. He put forth three basic desiderata, which form the basis for Bayesian probability theory:

1. Degrees of plausibility are represented by real numbers
2. The theory must have qualitative correspondence with common sense

3. The theory must be consistent.

From these three desiderata, the rules of probability theory, known as the sum rule and the product rule, can be derived. Given two hypotheses A and B and some information I , the sum rule tells one how to compute the probability for A OR B given that information I , which is represented by $p(A + B|I)$. Note that the “+” symbol inside the parentheses represents an OR operator. The sum rule takes the form

$$p(A + B|I) = p(A|I) + p(B|I) - p(AB|I). \quad (1)$$

It stipulates that the probability for A OR B given I is equal to the probability for A given only the information I , plus the probability for B given I , less the joint probability for A AND B given I . The product rule tells one how to form the probability for A AND B given the information I , which is represented by $p(AB|I)$. The product rule,

$$p(AB|I) = p(A|BI) p(B|I), \quad (2)$$

indicates that the probability for A AND B given I is equal to the probability for A given both B and I times the probability for B given only I . However, this factorization is not unique; the probability for B and A is the same as the probability for A and B , so one could also choose the factorization

$$p(AB|I) = p(B|AI) p(A|I). \quad (3)$$

Equating Eq. (2) and Eq. (3) gives

$$p(A|BI) p(B|I) = p(B|AI) p(A|I), \quad (4)$$

and dividing by $p(B|I)$ yields

$$p(A|BI) = \frac{p(B|AI) p(A|I)}{p(B|I)}, \quad (5)$$

which is Bayes' theorem. Bayes' theorem represents learning from experience; if one knows the probability for A given some information I , and subsequently acquires some new information B , then Bayes' theorem dictates how the new information should be incorporated to determine the probability for A given both B and I . Bayes' theorem becomes more intuitive with a change of notation. If the symbol A is replaced by some hypothesis H and B is replaced by some data D , then Bayes' theorem can be written

$$p(H|DI) = \frac{p(D|HI) p(H|I)}{p(D|I)}. \quad (6)$$

In this form, it can be seen that Bayes' theorem dictates how data should be incorporated to update one's existing knowledge regarding a hypothesis.

Each term in Bayes' theorem has a name. The term $p(H|DI)$ is called a posterior probability. It is the probability for a hypothesis when the data and all relevant background information

have been taken into account. The term $p(D|HI)$ is the direct probability for the data, and in some cases is proportional to a likelihood function. The term $p(H|I)$ is called the prior probability for the hypothesis. It allows the researcher to incorporate any information known about the hypothesis before data collection. The term in the denominator, $p(D|I)$, is called the evidence or the global likelihood, and it often serves as a normalization constant.

2.2. Structure of the Algorithm

The Bayesian calculations discussed below have been implemented in a computer program that identifies features in the EEG. The algorithm has been designed to identify smooth delta waves, delta brushes, interburst intervals, and general activity. It is divided into two steps, each requiring its own calculation for the posterior probability (see Appendixes A, B for details on obtaining the full calculations). The first step identifies delta waves, interburst intervals, and periods of general activity, where general activity is any section of the EEG recording not containing an interburst interval or delta wave. The second step takes the delta waves identified in the first step and attempts to further subdivide them by classifying them as either smooth delta waves or delta brushes. The algorithm was implemented in real-time using Fortran 95 running in parallel on an eight processor Linux PC.

2.3. Step One: Delta Waves, Interburst Intervals, and General Activity

The first step in the algorithm is to identify interburst intervals and delta waves. To achieve this identification, the algorithm must allocate the EEG data $\{d_1, d_2, \dots, d_m\} = D$ to one of three classes: interburst interval, delta wave, and general activity. The goal then is to calculate the posterior probability for each class $p(M|DI)$, where the classes are represented symbolically by M , and $M = 1$ designates the interburst interval class, $M = 2$ designates the delta wave class, and $M = 3$ designates the general activity class. Once the posterior probabilities for each class have been computed, the algorithm categorizes the data based on which class has the highest posterior probability.

To calculate the posterior probability for each class, the data must be related to the classes. Each class is comprised of one or more patterns, and each pattern is represented symbolically by a function which is designated by $\mathcal{P}_{jM}$, where $\mathcal{P}_{jM}$ is the j th pattern in class M . So, the equation that relates the data to the classes is of the form

$$d_t = \mathcal{P}_{jM}(t) + n_t, \quad (7)$$

where n_t is additive noise. Each pattern $\mathcal{P}_{jM}$ consists of a vector of N_m amplitudes, $\mathbf{A}_{jM}$, as well as the prior information related to those amplitudes. For simplicity, the notation

$$\mathcal{P}_{jM}(t) = A_{tjM} \quad (8)$$

is adopted with the understanding that the prior for the amplitude A_{tjM} is assigned using the prior information about the pattern $\mathcal{P}_{jM}(t)$. The prior information for each pattern in each class is unique. In particular, for an amplitude at time t , the prior information about the pattern specifies the most probable value of the amplitude, the low and high range of the amplitude, and the variability of the amplitude in a given pattern.

In the case of the interburst interval class, only one pattern is required, and the prior information about the interburst interval restricts the amplitudes to low values. More specifically, the prior information dictates that the most probable value of an amplitude in the interburst interval class is zero μV for the duration of the interval, and that the bounds on the amplitudes are $-10\mu V$ and $10\mu V$. The interburst interval pattern is shown in Fig. 2(a), with the solid line representing the most probable values of the amplitudes and the dotted lines representing the bounds on the amplitudes.

The general activity class also contains only one pattern, which is shown in Fig. 2(b). This pattern is similar to the interburst interval pattern, but the bounds used in this pattern are much larger than those used in the interburst interval pattern.

Finally, in the case of the delta wave class, multiple patterns are required to describe the various morphologies of delta waves. An example of a delta wave pattern is shown in Fig. 2(c). One can see that the solid line, which is the most probable value of the amplitudes, varies to reflect the shape of the delta wave. Additionally, the high and low bounds are $\pm 50\mu V$ from the most probable value to allow for variations in the shape of the delta waves. To demonstrate how the delta wave patterns differ from each other, several are plotted in Fig. 3.

2.4. Step Two: Delta Brushes and Smooth Delta Waves

After having identified delta waves, the algorithm must distinguish between smooth delta waves and delta brushes, and a separate calculation is required to do this. The smooth delta wave patterns and the delta brush patterns are identical in shape to the delta wave patterns used in the first step of the algorithm (see Fig. 3). However, the prior information regarding the amplitudes differs. First, the bounds on the amplitudes are eliminated because the feature in question has already been identified as a delta wave. Second, in order to distinguish between smooth delta waves and delta brushes, a smoothness constraint is placed on the amplitudes for a smooth delta wave pattern. For delta brush patterns, there is no such constraint.

2.5. EEG data

EEG recordings were obtained on 233 neonates at Royal Women's Hospital and Royal Children's Hospital in Melbourne, Australia between April 2001 and December 2003. The recordings were taken using a 2-channel BrainZ BRM2 monitor with a 0.1 Hz high-pass filter. Of the 233, 14 infants were selected for further study. These infants had no intraventricular hemorrhage and no abnormal cranial ultrasound studies throughout their course in the NICU, as well as normal mental (Mental Development Index > 85) and psychomotor development (Psychomotor Developmental Index > 85) at two years of age. The average gestational age of the 14 infants was 28.1 (27-29.6) weeks, and the average PMA at the time of recording was 29.4 (28-31.6) weeks. For each infant, a 10 minute epoch was selected for further study. These epochs were chosen based on their high number of delta brushes, smooth delta waves, and interburst intervals. The selected EEG epochs were required to have a maximum impedance of $10k\Omega$, but otherwise no artifact rejection was performed.

Two experienced electroencephalographers, both of whom were blinded to the results of the algorithm, were given the 14 recordings (140 minutes of data) and asked to independently identify interburst intervals, smooth delta waves, and delta brushes. In order to have consistent definitions of these waves, the electroencephalographers were asked to identify positive delta waves that were at least $100\mu V$ in amplitude and between 0.5 Hz and 1.5 Hz; for interburst intervals, they were asked to consider only those sections in which the amplitude remained under $10\mu V$ and exceeded 5 seconds. For consistency, only the left channel was studied for each infant. Two of the 14 recordings were used by the algorithm designer to modify the patterns to give the best agreement between the algorithm and the electroencephalographers. The remaining 12 recordings were used to compare the algorithm and the electroencephalographers, leaving a total of 2 hours for comparison.

3. Results

The results for the identification of delta waves and inter-burst intervals is shown in Tables 1-3. Table 1 offers a general comparison between the algorithm and both readers. From this table, it can be seen that the algorithm is conservative in its classification, identifying fewer delta waves and interburst intervals than either Reader 1 or Reader 2.

Table 2 offers a more direct comparison between the algorithm and both readers when identifying delta waves. In Table 2(a), Reader 1 is taken as the gold standard and both the algorithm and Reader 2 are compared with Reader 1. The algorithm correctly identified 557 of the 709 delta waves identified by Reader 1, yielding a true positive rate of 79%. However, of the 686 delta waves identified by the algorithm, 129 were not identified by Reader 1, giving a false positive rate of 19% for the algorithm. In order to accurately judge the algorithm's performance, these numbers must be compared with Reader 2's performance. From Table 2(a), Reader 2 had a true positive rate of 82% with a false positive rate of 23%, meaning that the algorithm has both a slightly lower true positive rate and slightly lower false positive rate when compared with Reader 2. This trend continues when Reader 2 is taken as the gold standard for identifying delta waves, which is shown in Table 2(b).

Table 3 displays the results of the algorithm's performance when identifying interburst intervals. When Reader 1 is taken as the gold standard, Table 3(a), the algorithm correctly identified 72% of interburst intervals while having a false positive rate of 20%. For comparison, Reader 2 identified 74% interburst intervals correctly with 23% of the interburst intervals identified by Reader 2 being false positives. Similar results are found in Table 3(b), in which Reader 2 is taken as the gold standard. Just as in the selection of delta waves, the algorithm consistently identified interburst intervals at a lower true positive rate than the electroencephalographers, but also maintained a lower false positive rate.

Table 4 contains the results of the classification of delta waves into smooth delta waves and delta brushes. Only delta waves identified by both readers and the algorithm were used for this comparison. The algorithm did not perform as well at distinguishing between smooth delta waves and delta brushes as it did when identifying the delta waves, which can be seen in table 4(a), where Reader 1 was taken as the gold standard. In this case, the algorithm identified 87% of smooth delta waves identified by Reader 1, while correctly identifying

53% of the delta brushes identified by Reader 1. Reader 2 identified 89% of smooth delta waves identified by Reader 1 while correctly identifying 67% of delta brushes. The algorithm also had lower true positive rates when Reader 2 was taken as the gold standard (see Table 4(b)).

The algorithm can also be evaluated by focusing on the classifications in which both electroencephalographers are in agreement (Table 5). In those instances, the algorithm performs well when identifying interburst intervals, delta waves, and smooth delta waves; for these features, the true positive rates are all at 86% or higher with false positive rates at 12% or lower. For the identification of delta brushes, however, the true positive rate is substantially lower at 61% with a higher false positive rate of 46%.

4. Discussion

The performance of the algorithm must be evaluated relative to the inter-rater reliability. In the case of the identification of delta waves and interburst intervals, the algorithm was in agreement with the two electroencephalographers nearly as much as they were in agreement with each other. The algorithm was designed to maximize the true positive rate while maintaining a low false positive rate; because the true positive detection rates were only a few percent lower than the readers', the algorithm's performance was comparable to the readers.

When distinguishing between smooth delta waves and delta brushes, the agreement between the electroencephalographers and the algorithm was not as strong. In this case, the readers were in better agreement with each other than the algorithm was with either of them. Of particular concern is that 46% of the delta brushes identified by the algorithm were classified as smooth delta waves by both Reader 1 and Reader 2 (see Table 5). While this number is high, it should be noted that 34% of the delta brushes identified by Reader 2 were classified as smooth delta waves by both Reader 1 and the algorithm. A similar cause for concern, as shown in Table 4(a), is that the algorithm identified more false positive delta brushes ($432-375=57$) than it did true positives (35). However, this result is simply the result of the data containing an order of magnitude more smooth delta waves than delta brushes. Indeed, Table 4(a) similarly shows that Reader 2 also identified more false positive delta brushes at 46 than true positive delta brushes at 44. These results underscore the difficulty in identifying delta brushes. There exists a continuum of waves between perfectly smooth delta waves and "classic" delta brushes, so forcing them to be classified into one of two categories will inevitably lead to disagreements, even between experienced electroencephalographers. Ultimately, the results of these markings indicate that while the algorithm does not perform as well as the expert readers, its performance is reasonable given the level of disagreement that exists between the expert readers themselves.

Having demonstrated that the algorithm can identify developmental landmarks nearly as consistently as experienced electroencephalographers, there are two main advantages to its use. The first advantage is consistency. While the algorithm may not always agree with the electroencephalographers, it will make its classifications consistently, and having consistent measures may prove useful in correlating EEG development with outcome. Second, the

algorithm has been designed to run in real-time, thus providing the physician with immediate feedback regarding the state of the infant.

Despite these advantages, there algorithm has a number of weaknesses as currently implemented. First, because the algorithm is focused on identifying specific developmental landmarks and these landmarks represent a small percentage of the EEG recording, the algorithm presently ignores much of the EEG recording. If these portions of the recording contain waves indicative of abnormal neurologic development, the algorithm will be ignorant of this. This shortcoming can be overcome to some extent with the addition of more developmental landmarks, but because much of the typical premature EEG recording contains waves of nonspecific morphology, much of the recording will remain unclassified by the algorithm. Second, the algorithm has no method for automatic artifact detection. If it is to be implemented in the NICU, it will have to rely on its users to identify portions of the EEG recording that have been compromised. Third, as with any automated technique, the algorithm's results will be complicated by the use of sedatives.

The current algorithm is a simple proof of principle, and for this reason its focus has been the identification of a limited number of developmental landmarks. However, the algorithm has the potential to be more powerful than its current implementation. It can be expanded to identify other developmental landmarks of interest through the creation of additional patterns. Furthermore, if implemented on a multiple-channel EEG recording, the algorithm will be capable of offering quantitative measures of synchrony of synchrony by comparing the results from different channels.

5. Conclusions

EEG offers a noninvasive and continuous method for monitoring the premature brain, but its potential has yet to be fully realized because of the subjectivity associated with its interpretation, a lack of standardized measures, a lack of EEG readers, and the delay between recording and diagnosis. These shortcomings can be addressed with an automated approach to EEG analysis that provides real-time, objective information to clinicians. This project has used Bayesian probability theory to consistently identify developmental landmarks in the EEG recording, and thus provide a first step towards such an approach. Its flexibility allows for additional developmental landmarks to be identified, which leaves open the possibility to gain further insight into the developing brain with future work.

Acknowledgments

This research was supported by grants from the National Institute of Health (R01HD057098 and P30HD062171) and the Doris Duke Distinguished Clinical Scientist Award.

7. Appendix

The software is available for others. For inquiries regarding the use of the algorithm, contact the corresponding author. The full Bayesian calculations associated with this article can be found below in Appendixes A, B.

Appendix A. Step One Calculation

The goal, when given an EEG recording containing a sequence of data values $\{d_1, d_2, \dots, d_m\} = D$, is to determine whether the data contain a delta wave, interburst interval, or general activity. If M represents either the delta wave, interburst interval, or general activity class, then to identify the class present in the data, one computes the posterior probability for each class given the data and the prior information, $p(M|DI)$. This posterior probability is given by Bayes' theorem

$$p(M|DI) = \frac{p(D|MI) p(M|I)}{p(D|I)}, \quad (\text{A.1})$$

where $p(D|MI)$ is the direct probability for the data given the class, $p(D|I)$ is the direct probability for the data given only the prior information, and $p(M|I)$ is the probability for the class given only the prior information, which is sometimes called a prior probability. The denominator in this equation is a normalization constant. To see this, one reintroduces the class M into $p(D|I)$, which yields

$$p(M|DI) = \frac{p(D|MI) p(M|I)}{\sum_{M=1}^3 p(DM|I)}, \quad (\text{A.2})$$

where there are 3 classes to be summed over in the denominator. Applying the product rule to the denominator yields

$$p(M|DI) = \frac{p(D|MI) p(M|I)}{\sum_{M=1}^3 p(D|MI) p(M|I)}. \quad (\text{A.3})$$

In this form, it is easily seen that the denominator is simply a normalization constant, and so it is omitted with the understanding that the posterior probability must be normalized at the end of calculation. The posterior probability for the class M then becomes

$$p(M|DI) \propto p(D|MI) p(M|I). \quad (\text{A.4})$$

In this calculation it is assumed that each class is equally probable, so a uniform prior probability is assigned to $p(M|I)$, and this constant can be included in the proportionality to obtain

$$p(M|DI) \propto p(D|MI). \quad (\text{A.5})$$

In general, the class M contains multiple patterns, and these patterns are introduced through the sum rule

$$p(M|DI) \propto \sum_{j=1}^{N_M} p(D \mathcal{P}_{jM} | MI), \quad (\text{A.6})$$

where there are N_M patterns in class M and $\mathcal{P}_{jM}$ is the j th pattern contained in class M .

Applying the product rule to factor $p(D \mathcal{P}_{jM} | MI)$ yields

$$p(M|DI) \propto \sum_{j=1}^{N_M} p(D|\mathcal{P}_{jM}MI) p(\mathcal{P}_{jM}|MI). \quad (\text{A.7})$$

Like the prior probability for the class, a uniform prior probability is assigned to the patterns. The equation for the posterior probability for class M then simplifies to

$$p(M|DI) \propto \frac{1}{N_M} \sum_{j=1}^{N_M} p(D|\mathcal{P}_{jM}MI). \quad (\text{A.8})$$

Unlike the denominator $p(D|I)$ and the prior probability for the models $p(M|I)$, the $1/N_M$ term that comes from $p(\mathcal{P}_{jM}|MI)$ can not be included in the proportionality constant because the number of patterns in a class is class-dependent.

The probability for a particular class M is equal to a summation of the direct probabilities for the data given the patterns contained in that class. However, the calculation for the direct probability for the data given any pattern is identical, so for notational simplicity, the calculation is done using a single generic pattern with the understanding that the notation must be revised to include things like the pattern number and the class at its completion. In the calculation that follows, this generic pattern will be represented by $\mathcal{P}$, and the amplitudes associated with this pattern are designated as $\{A_1, A_2, \dots, A_m\} = \mathbf{A}$. To proceed with the calculation of the direct probability for the data given a generic pattern, $p(D|\mathcal{P}I)$, the amplitudes $\mathbf{A}$ of the pattern $\mathcal{P}$ are introduced by application of the sum rule

$$p(D|\mathcal{P}I) = \int p(D\mathbf{A}|\mathcal{P}I) d\mathbf{A}. \quad (\text{A.9})$$

This is an m -dimensional integral, where m is the total number of amplitudes. Using the product rule to factor $p(D\mathbf{A}|\mathcal{P}I)$, Eq. (A.9) becomes

$$p(D|\mathcal{P}I) = \int p(D|\mathbf{A}\mathcal{P}I) p(\mathbf{A}|\mathcal{P}I) d\mathbf{A}. \quad (\text{A.10})$$

The product rule can be used to factor the prior probabilities for the amplitudes. Applying the product rule to the right-hand term in Eq. (A.10), one obtains

$$p(\mathbf{A}|\mathcal{P}I) = p(A_2 \dots A_m | A_1 \mathcal{P}I) p(A_1 | \mathcal{P}I). \quad (\text{A.11})$$

If the prior probabilities for the amplitudes are assumed to be logically independent, Eq. (A.11) simplifies to

$$p(\mathbf{A}|\mathcal{P}I) = p(A_2 \dots A_m | \mathcal{P}I) p(A_1 | \mathcal{P}I), \quad (\text{A.12})$$

and repeated application of the product rule yields

$$p(D|\mathcal{P}I) = \int p(D|\mathbf{A}\mathcal{P}I) \prod_{t=1}^m [p(A_t | \mathcal{P}I)] d\mathbf{A}. \quad (\text{A.13})$$

At this point the likelihood and the prior probabilities for the amplitudes must be assigned. A Gaussian is assigned to the likelihood, which takes the form

$$p(D|\mathbf{A}, \mathcal{P}I) = \prod_{t=1}^m \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{-\frac{(d_t - A_t)^2}{2\sigma^2}\right\}, \quad (\text{A.14})$$

where σ is the standard deviation of the noise prior probability. The patterns are used to construct the prior probabilities for the amplitudes, which are represented by bounded Gaussians

$$p(A_t|\mathcal{P}I) \approx \begin{cases} \frac{1}{\sqrt{2\pi}\beta_t^2} \exp\left\{-\frac{(A_t - u_t)^2}{2\beta_t^2}\right\} & \text{if } L_t \leq A_t \leq H_t \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A.15})$$

where u_t is the most probable value of the amplitude A_t , β_t is the standard deviation of the Gaussian prior probability, and L_t and H_t represent the low and high bounds at time t . Equation (A.15) is an approximation because the Gaussian is bounded by L_t and H_t . For this approximation to be valid, the bounds must not significantly truncate the Gaussian prior. In this study, the approximation is good to within one part in 10^4 .

Taking Eq. (A.15) as an equality and substituting it and the likelihood from Eq. (A.14) into Eq. (A.10), the direct probability for the data becomes

$$p(D|\mathcal{P}I) = \prod_{t=1}^m \int_{L_t}^{H_t} (2\pi\beta_t\sigma)^{-1} \exp\left\{-\frac{(A_t - d_t)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{(A_t - u_t)^2}{2\beta_t^2}\right\} dA_t. \quad (\text{A.16})$$

When evaluating this integral, it is helpful both notationally and conceptually to introduce the term $\hat{A}_t$, which is the expected value of the amplitude A_t and defined as

$$\hat{A}_t \equiv \frac{\beta_t^2 d_t + \sigma^2 u_t}{\beta_t^2 + \sigma^2}. \quad (\text{A.17})$$

Rewriting Eq. (A.16) in terms of the expected amplitudes $\hat{A}_t$, the direct probability for the data becomes

$$p(D|\mathcal{P}I) = \prod_{t=1}^m (2\pi\beta_t\sigma)^{-1} \exp\left\{-\frac{[\beta_t^2 d_t^2 + \sigma^2 u_t^2]}{2\beta_t^2 \sigma^2}\right\} \times \int_{L_t}^{H_t} \exp\left\{-\frac{(\beta_t^2 + \sigma^2)}{2\beta_t^2 \sigma^2} (A_t^2 - 2A_t \hat{A}_t)\right\} dA_t. \quad (\text{A.18})$$

Completing the square

$$p(D|\mathcal{P}I) = \prod_{t=1}^m (2\pi\beta_t\sigma)^{-1} \exp\left\{-\frac{[\beta_t^2 d_t^2 + \sigma^2 u_t^2]}{2\beta_t^2 \sigma^2}\right\} \times \exp\left\{+\frac{(\beta_t^2 + \sigma^2)}{2\beta_t^2 \sigma^2} \hat{A}_t^2\right\} \times \int_{L_t}^{H_t} \exp\left\{-\frac{(\beta_t^2 + \sigma^2)}{2\beta_t^2 \sigma^2} (A_t - \hat{A}_t)^2\right\} dA_t \quad (\text{A.19})$$

and changing variables

$$\lambda_t^2 \equiv \frac{\beta_t^2 + \sigma^2}{2\beta_t^2 \sigma^2} (A_t - \hat{A}_t)^2 \quad (\text{A.20})$$

so that

$$dA_t = \sqrt{\frac{2\beta_t^2\sigma^2}{\beta_t^2 + \sigma^2}} d\lambda_t, \quad (\text{A.21})$$

gives

$$\int_{L'_t}^{H'_t} e^{-\lambda_t^2} d\lambda_t = \frac{\sqrt{\pi}}{2} \left\{ \text{erfc}(L'_t) - \text{erfc}(H'_t) \right\}, \quad (\text{A.22})$$

where

$$L'_t \equiv \frac{\sqrt{\beta_t^2 + \sigma^2}}{\sqrt{2}\beta_t\sigma} [L_t - \hat{A}_t] \quad (\text{A.23})$$

and

$$H'_t \equiv \frac{\sqrt{\beta_t^2 + \sigma^2}}{\sqrt{2}\beta_t\sigma} [H_t - \hat{A}_t]. \quad (\text{A.24})$$

Combining Eqs. (A.17) and (A.19-A.22) and simplifying the result yields

$$p(D|\mathcal{D}I) = \prod_{t=1}^m \frac{[\text{erfc}(L'_t) - \text{erfc}(H'_t)]}{2\sqrt{2\pi(\beta_t^2 + \sigma^2)}} \exp \left\{ \frac{-(d_t - u_t)^2}{2(\beta_t^2 + \sigma^2)} \right\}. \quad (\text{A.25})$$

This is the final result for the direct probability of the data given a generic pattern.

To obtain the posterior probability for a class, Eq. (A.25) must be substituted into Eq. (A.8).

With a change in notation to represent a specific pattern rather than the generic pattern, the posterior probability for a class becomes

$$p(M|DI) \propto \frac{1}{N_M} \sum_{j=1}^{N_M} \prod_{t=1}^m \frac{[\text{erfc}(L'_{tjM}) - \text{erfc}(H'_{tjM})]}{2\sqrt{2\pi(\beta_{tjM}^2 + \sigma^2)}} \times \exp \left\{ \frac{-(d_t - u_{tjM})^2}{2(\beta_{tjM}^2 + \sigma^2)} \right\}, \quad (\text{A.26})$$

where the subscripts tjM designate the time, the pattern, and the class respectively.

Additionally, L'_{tjM} , H'_{tjM} , and $\hat{A}_{tjM}$ are given by

$$L'_{tjM} \equiv \frac{\sqrt{\beta_{tjM}^2 + \sigma^2}}{\sqrt{2}\beta_{tjM}\sigma} [L_{tjM} - \hat{A}_{tjM}], \quad (\text{A.27})$$

$$H'_{tjM} \equiv \frac{\sqrt{\beta_{tjM}^2 + \sigma^2}}{\sqrt{2}\beta_{tjM}\sigma} [H_{tjM} - \hat{A}_{tjM}] \quad (\text{A.28})$$

and

$$\hat{A}_{tjM} \equiv \frac{\beta_{tjM}^2 d_t + \sigma^2 u_{tjM}}{\sigma^2 + \beta_{tjM}^2}. \quad (\text{A.29})$$

Appendix B. Step Two Calculation

The method for classifying smooth delta waves and delta brushes is similar to the method taken in Appendix A. When given a sequence of data $\{d_1, d_2, \dots, d_m\} = D$ from an EEG recording, the goal is to evaluate the posterior probability $p(M|DI)$ for the classes of interest M . However, in this portion of the algorithm, the data must be classified as either a smooth delta wave or a delta brush, so $M = 1$ will designate the smooth delta wave class and $M = 2$ will designate the delta brush class. Each class contains multiple patterns, which are represented symbolically by $\mathcal{P}_{jM}$, where $\mathcal{P}_{jM}$ is the j th pattern in class M . The equation relating the data to each class is given by

$$d_t = \mathcal{P}_{jM}(t) + n_t, \quad (\text{B.1})$$

where n_t is an additive noise. Again, the shorthand notation

$$\mathcal{P}_{jM}(t) = A_{tjM} \quad (\text{B.2})$$

is adopted with the understanding that the prior for the amplitude A_{tjM} is given by $\mathcal{P}_{jM}(t)$.

The posterior probability for a class, $p(M|DI)$, is given by Bayes' theorem

$$p(M|DI) = \frac{p(D|MI)p(M|I)}{p(D|I)}. \quad (\text{B.3})$$

The calculation follows the same steps as shown in Eqs. (A.1-A.9), so that the equation for the direct probability for the data, given a generic pattern, is given by

$$p(D|\mathcal{P}I) = \int p(D\mathbf{A}|\mathcal{P}I) d\mathbf{A}. \quad (\text{B.4})$$

This is an m -dimensional integral, where m represents the total time of the pattern. Note that there are no bounds on the amplitudes, so the limits of the integration extend to infinite. Additionally, unlike the previous calculation, there is prior information regarding the numerical values of adjacent amplitudes. Consequently, the prior probabilities for the amplitudes are not logically independent, and care must be taken to factor the priors correctly. The product rule is used to factor A_1 , A_2 , and A_3

$$\begin{aligned} p(D|\mathcal{P}I) &= \int p(DA_2 \dots A_m | A_1 \mathcal{P}I) p(A_1 | \mathcal{P}I) d\mathbf{A}. \\ &= \int p(DA_3 \dots A_m | A_1 A_2 \mathcal{P}I) p(A_1 | \mathcal{P}I) p(A_2 | A_1 \mathcal{P}I) d\mathbf{A}. \\ &= \int p(DA_4 \dots A_m | A_1 A_2 A_3 \mathcal{P}I) p(A_1 | \mathcal{P}I) p(A_2 | A_1 \mathcal{P}I) \times p(A_3 | A_1 A_2 \mathcal{P}I) d\mathbf{A}. \end{aligned} \quad (\text{B.5})$$

At this point the pattern can be seen, and factoring all amplitudes yields

$$p(D|\mathcal{P}I) = \int p(D|\mathbf{A}, \mathcal{P}I) p(A_1|\mathcal{P}I) \times \prod_{t=2}^m p(A_t|A_{t-1} \dots A_1 \mathcal{P}I) d\mathbf{A}, \quad (\text{B.6})$$

where the prior probability for A_1 has been written separately because it does not have a dependence on any other amplitudes. The probabilities are then

$$p(D|A_1 \dots A_m \mathcal{P}I) = \prod_{t=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(A_t - d_t)^2}{2\sigma^2} \right\} \quad (\text{B.7})$$

for the likelihood,

$$p(A_1|\mathcal{P}I) = k_1 \exp \left\{ \frac{-(A_1 - u_1)^2}{2\beta_1^2} \right\} \quad (\text{B.8})$$

for the prior probability for A_1 , and

$$p(A_t|A_{t-1} \dots A_1 \mathcal{P}I) = k_2 \exp \left\{ \frac{-(A_t - u_t)^2}{2\beta_t^2} \right\} \times \exp \left\{ \frac{-(A_t - A_{t-1})^2}{2\gamma_t^2} \right\} \quad (\text{B.9})$$

for the prior probability for A_t given amplitudes A_1 through A_{t-1} . Equation (B.9) contains a hyperparameter γ_t , which imposes the smoothness constraint. For delta wave patterns, a small value of γ_t requires that consecutive amplitudes be close in magnitude (smooth). However, for delta brush patterns, a large value of γ_t imposes no smoothness constraint and leaves the delta brush patterns free to model brushes when they are present. The constants k_1 and k_2 are required to ensure the probabilities are properly normalized, but the normalization is complicated and is computed in Appendix B.1. For now, the calculation will continue with $k = k_1 k_2$ representing the normalization constant. Substituting Eqs. (B.7-B.9) into Eq. (B.6) gives

$$p(D|\mathcal{P}I) = k \int \prod_{t=1}^m \left[(2\pi\sigma^2)^{-1/2} \exp \left\{ \frac{-(A_t - d_t)^2}{2\sigma^2} \right\} \right] \times \exp \left\{ \frac{-(A_1 - u_1)^2}{2\beta_1^2} \right\} \times \prod_{t=2}^m \left[\exp \left\{ \frac{-(A_t - u_t)^2}{2\beta_t^2} \right\} \exp \left\{ \frac{-(A_t - A_{t-1})^2}{2\gamma_t^2} \right\} \right] d\mathbf{A}, \quad (\text{B.10})$$

as the direct probability for the data. Separating out $t = 1$ from the product and rearranging yields

$$\begin{aligned} p(D|\mathcal{P}I) &= k \left[(2\pi\sigma^2)^{-m/2} \right] \\ &\times \int \exp \left\{ \frac{-(A_1 - u_1)^2}{2\beta_1^2} + \frac{-(A_1 - d_1)^2}{2\sigma^2} \right\} \\ &\times \exp \left\{ \sum_{t=2}^m \left[\frac{-(A_t - d_t)^2}{2\sigma^2} + \right] \right\} \\ &\times \exp \left\{ \sum_{t=2}^m \left[\frac{-(A_t - u_t)^2}{2\beta_t^2} + \frac{-(A_t - A_{t-1})^2}{2\gamma_t^2} \right] \right\} d\mathbf{A} \end{aligned} \quad (\text{B.11})$$

for the direct probability. Gathering like terms, Eq. (B.11) can be rewritten in the form

$$p(D|\mathcal{P}I) = k \left[(2\pi\sigma^2)^{-m/2} \right] \exp \left\{ -\frac{1}{2} \sum_{t=1}^m \left[\frac{d_t^2}{\sigma^2} + \frac{u_t^2}{\beta_t^2} \right] \right\} \times \int \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^m \sum_{k=1}^m A_j R_{jk} A_k - 2 \sum_j A_j S_j \right] \right\} d\mathbf{A}, \quad (\text{B.12})$$

where the elements of the R-matrix are given by

$$\left\{ \begin{array}{ll} R_{11} = \sigma^{-2} + \beta_1^{-2} + \gamma_1^{-2} \\ R_{mm} = \sigma^{-2} + \beta_m^{-2} + \gamma_m^{-2} \\ R_{jj} = \sigma^{-2} + \beta_j^{-2} + 2\gamma_j^{-2} & 1 < j < m \\ R_{j-1,j} = -\gamma_j^{-2} & 1 < j \leq m \\ R_{j,j-1} = -\gamma_j^{-2} & 1 < j \leq m \\ R_{jk} = 0 & \text{otherwise} \end{array} \right\} \quad (\text{B.13})$$

and the elements of the S-vector are given by

$$S_j = d_j \sigma^{-2} + u_j \beta_j^{-2} \quad 1 \leq j \leq m. \quad (\text{B.14})$$

Defining a function $f(\mathbf{A})$,

$$f(\mathbf{A}) \equiv \frac{1}{2} \mathbf{A}^T \mathbf{R} \mathbf{A} - \mathbf{A} \mathbf{S}, \quad (\text{B.15})$$

and Taylor expanding about the amplitudes $\hat{\mathbf{A}}$ that maximize $f(\mathbf{A})$, one obtains

$$f(\mathbf{A}) \approx f(\hat{\mathbf{A}}) + \frac{1}{2} (\mathbf{A} - \hat{\mathbf{A}})^T \frac{\partial^2}{\partial \mathbf{A}^2} f(\mathbf{A}) (\mathbf{A} - \hat{\mathbf{A}}), \quad (\text{B.16})$$

where

$$f(\hat{\mathbf{A}}) = \frac{1}{2} \hat{\mathbf{A}}^T \mathbf{R} \hat{\mathbf{A}} - \hat{\mathbf{A}} \mathbf{S} \quad (\text{B.17})$$

and

$$\frac{\partial^2}{\partial \mathbf{A}^2} f(\mathbf{A}) = \mathbf{R}. \quad (\text{B.18})$$

Substituting Eqs. (B.17) and (B.18) into Eq. (B.16) yields

$$f(\mathbf{A}) = \frac{1}{2} \hat{\mathbf{A}}^T \mathbf{R} \hat{\mathbf{A}} - \hat{\mathbf{A}} \mathbf{S} + \frac{1}{2} (\mathbf{A} - \hat{\mathbf{A}})^T \mathbf{R} (\mathbf{A} - \hat{\mathbf{A}}). \quad (\text{B.19})$$

where the peak amplitudes $\hat{\mathbf{A}}$ are given by

$$\hat{\mathbf{A}} = \mathbf{R}^{-1} \mathbf{S}. \quad (\text{B.20})$$

Note that the approximation in Eq. (B.15) has been replaced with an equality in Eq. (B.19) because $f(\mathbf{A})$ is quadratic in $\mathbf{A}$. Substituting Eqs. (B.15) and (B.19) into Eq. (B.12) gives

$$\begin{aligned}
p(D|\mathcal{P}I) &= k \left[(2\pi\sigma^2)^{-m/2} \right] \exp \left\{ -\frac{1}{2} \sum_{t=1}^m \left[\frac{d_t^2}{\sigma^2} + \frac{u_t^2}{\beta_t^2} \right] \right\} \\
&\times \exp \left\{ -\frac{1}{2} \hat{\mathbf{A}}^T \mathbf{R} \hat{\mathbf{A}} + \hat{\mathbf{A}} \mathbf{S} \right\} \\
&\times \int \exp \left\{ -\frac{1}{2} (\mathbf{A} - \hat{\mathbf{A}})^T \mathbf{R} (\mathbf{A} - \hat{\mathbf{A}}) \right\} d\mathbf{A}
\end{aligned} \quad (\text{B.21})$$

for the direct probability for the data. The integral in Eq. (B.21) is a multivariate Gaussian quadrature integral, which can be easily evaluated. See Sivya and Skilling [2], and Bretthorst [1] for details. Evaluating the integrals gives

$$\int \exp \left\{ -\frac{1}{2} (\mathbf{A} - \hat{\mathbf{A}})^T \mathbf{R} (\mathbf{A} - \hat{\mathbf{A}}) \right\} d\mathbf{A} = \frac{(2\pi)^{m/2}}{\sqrt{\det(\mathbf{R})}}, \quad (\text{B.22})$$

and substituting Eq. (B.22) into Eq. (B.21) yields

$$p(D|\mathcal{P}I) = \frac{k}{\sigma^{m/2} \sqrt{\det(\mathbf{R})}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^m \left[\frac{d_t^2}{\sigma^2} + \frac{u_t^2}{\beta_t^2} \right] \right\} \times \exp \left\{ -\frac{1}{2} \hat{\mathbf{A}}^T \mathbf{R} \hat{\mathbf{A}} + \hat{\mathbf{A}} \mathbf{S} \right\}. \quad (\text{B.23})$$

Equation (B.23) can be rewritten in summation notation, and takes a form similar to Eq. (B.11), where the integral has been evaluated. This change of notation gives

$$\begin{aligned}
p(D|\mathcal{P}I) &= \frac{k}{\sigma^{m/2} \sqrt{\det(\mathbf{R})}} \\
&\times \exp \left\{ \frac{-(\hat{A}_1 - u_1)^2}{2\beta_1^2} + \frac{-(\hat{A}_1 - d_1)^2}{2\sigma^2} \right\} \\
&\times \exp \left\{ \sum_{t=2}^m \left[\frac{-(\hat{A}_t - d_t)^2}{2\sigma^2} + \frac{-(\hat{A}_t - u_t)^2}{2\beta_t^2} + \frac{-(\hat{A}_t - \hat{A}_{t-1})^2}{2\gamma_t^2} \right] \right\},
\end{aligned} \quad (\text{B.24})$$

which is the direct probability for the data given a pattern. However, the normalization constant still needs to be evaluated, and this is the subject of the next subsection.

Appendix B.1. The Normalization Constant

To properly normalize the prior probabilities for the amplitudes from Eq. (B.6), the condition

$$1 = \int p(A_1|\mathcal{P}I) \prod_{t=2}^m p(A_t|A_{t-1} \dots A_1 \mathcal{P}I) d\mathbf{A} \quad (\text{B.25})$$

must be satisfied. Substituting Eqs. (B.8) and (B.9), and writing the product of exponentials as the exponential of a sum gives

$$k^{-1} = \int \exp \left\{ \frac{-(A_1 - u_1)^2}{2\beta_1^2} \right\} \times \exp \left\{ \sum_{t=2}^m \left[\frac{-(A_t - u_t)^2}{2\beta_t^2} + \frac{-(A_t - A_{t-1})^2}{2\gamma_t^2} \right] \right\} d\mathbf{A}, \quad (\text{B.26})$$

where $k = k_1 k_2$. Equation (B.26) is of the same form as Eq. (B.11), so the same steps are taken to evaluate the integral. Equation (B.26) is then rewritten in matrix form

$$k^{-1} = \exp \left\{ -\frac{1}{2} \sum_{t=1}^m \left[\frac{u_t^2}{\beta_t^2} \right] \right\} \int \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^m \sum_{k=1}^m A_j Z_{jk} A_k - 2 \sum_j A_j Y_j \right] \right\} d\mathbf{A}, \quad (\text{B.27})$$

where the elements of the $\mathbf{Z}$ matrix are given by

$$\left\{ \begin{array}{ll} Z_{11} = \beta_1^{-2} + \gamma_1^{-2} \\ Z_{mm} = \beta_m^{-2} + \gamma_m^{-2} \\ Z_{jj} = \beta_j^{-2} + 2\gamma_j^{-2} & 1 < j < m \\ Z_{j-1,j} = -\gamma_j^{-2} & 1 < j \leq m \\ Z_{j,j-1} = -\gamma_j^{-2} & 1 < j \leq m \\ Z_{jk} = 0 & \text{otherwise} \end{array} \right\} \quad (\text{B.28})$$

and the elements of the $\mathbf{Y}$ vector are

$$Y_j = u_j \beta_j^{-2} \quad 1 \leq j \leq m. \quad (\text{B.29})$$

In this calculation, the $\mathbf{Z}$ matrix is analogous to the $\mathbf{R}$ matrix defined in Eq. (B.13), and the $\mathbf{Y}$ vector is analogous to the $\mathbf{S}$ vector defined in Eq. (B.14). Though the details are omitted here, the previous calculation can be repeated using these new quantities, which yields

$$k = \frac{\sqrt{\det(\mathbf{Z})}}{(2\pi)^{m/2}} \exp \left\{ -\frac{(\hat{A}'_1 - u_1)^2}{2\beta_1^2} \right\} \times \exp \left\{ \sum_{t=2}^m \left[\frac{(\hat{A}'_t - u_t)^2}{2\beta_t^2} + \frac{(\hat{A}'_t - \hat{A}'_{t-1})^2}{2\gamma_t^2} \right] \right\} \quad (\text{B.30})$$

for the normalization constant, where $\hat{\mathbf{A}}'$ is the vector of peak amplitudes given by the equation $\hat{\mathbf{A}}' = \mathbf{Z}^{-1}\mathbf{Y}$. Combining Eqs. (B.24) and (B.30) yields the direct probability for the data given a general pattern

$$\begin{aligned} p(D|\mathcal{P}I) &= \frac{1}{(2\pi\sigma^2)^{m/2}} \frac{\sqrt{\det(\mathbf{Z})}}{\sqrt{\det(\mathbf{R})}} \\ &\times \exp \left\{ -\frac{(\hat{A}_1 - u_1)^2}{2\beta_1^2} + \frac{(\hat{A}_1 - d_1)^2}{2\sigma^2} + \frac{(\hat{A}'_1 - u_1)^2}{2\beta_1^2} \right\} \\ &\times \exp \left\{ \sum_{t=2}^m \left[\frac{-(\hat{A}_t - d_t)^2}{2\sigma^2} + \frac{-(\hat{A}_t - u_t)^2}{2\beta_t^2} + \frac{-(\hat{A}_t - \hat{A}_{t-1})^2}{2\gamma_t^2} \right] \right\} \quad (\text{B.31}) \\ &\times \exp \left\{ \sum_{t=2}^m \left[\frac{(\hat{A}'_t - u_t)^2}{2\beta_t^2} + \frac{(\hat{A}'_t - \hat{A}'_{t-1})^2}{2\gamma_t^2} \right] \right\}. \end{aligned}$$

To obtain the probability for a class given the data and background information, Eq. (B.31) is substituted into Eq. (A.8), yielding

$$\begin{aligned}
p(M|DI) \propto & \sum_{j=1}^N \frac{1}{(2\pi\sigma^2)^{m/2}} \frac{\sqrt{\det(\mathbf{Z})}}{\sqrt{\det(\mathbf{R})}} \\
& \times \exp \left\{ \frac{-\left(\hat{A}_{1jM} - u_{1jM}\right)^2}{2\beta_{1jM}^2} + \frac{-\left(\hat{A}_{1jM} - d_1\right)^2}{2\sigma^2} \right\} \\
& \times \exp \left\{ \frac{\left(\hat{A}'_{1jM} - u_{1jM}\right)^2}{2\beta_{1jM}^2} \right\} \\
& \times \exp \left\{ \sum_{t=2}^m \left[\frac{-\left(\hat{A}_{tjM} - d_t\right)^2}{2\sigma^2} + \frac{-\left(\hat{A}_{tjM} - u_{tjM}\right)^2}{2\beta_{tjM}^2} \right] \right\} \\
& \times \exp \left\{ \sum_{t=2}^m \left[\frac{-\left(\hat{A}_{tjM} - \hat{A}_{t-1jM}\right)^2}{2\gamma_{tjM}^2} \right] \right\} \\
& \times \exp \left\{ \sum_{t=2}^m \left[\frac{\left(\hat{A}'_{tjM} - u_{tjM}\right)^2}{2\beta_{tjM}^2} + \frac{\left(\hat{A}'_{tjM} - \hat{A}'_{t-1jM}\right)^2}{2\gamma_{tjM}^2} \right] \right\}, \tag{B.32}
\end{aligned}$$

which is the posterior probability for a class M given the EEG data D and background information I , and is the equation implemented in the algorithm to determine if a delta wave is a smooth delta wave or a delta brush.

References

- Altenburg J, Vermeulen RJ, Strijers RLM, Fetter WPF, Stam CJ. Seizure detection in the neonatal EEG with synchronization likelihood. *Clin Neurophysiol.* 2003; 114:50–55. [PubMed: 12495763]
- Bretthorst, GL. *Lecture Notes in Statistics.* Springer-Verlag; Berlin: 1988.
- Cox RT. Probability, Frequency and Reasonable Expectation. *Am J Phys.* 1946; 14:1–13.
- Deburchgraeve W, Cherian PJ, De Vos M, Swarte RM, Blok JH, Visser GH, et al. Automated neonatal seizure detection mimicking a human observer reading EEG. *Clin Neurophysiol.* 2008; 119:2447–2454. [PubMed: 18824405]
- Eiselt M, Schendel M, Witte H, Dörschel J, Curzi-Dascalova L, D'Allest AM, et al. Quantitative analysis of discontinuous EEG in premature and full-term newborns during quiet sleep. *Electroencephalogr Clin Neurophysiol.* 1997; 103:528–34. [PubMed: 9402883]
- Faust O, Acharya UR, Min LC, Spath BHC. Automatic identification of epileptic and background EEG signals using frequency domain parameters. *Int J Neural Syst.* 2010; 20:159–176. [PubMed: 20411598]
- Gotman J, Flanagan D, Zhang J, Rosenblatt B. Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalogr Clin Neurophysiol.* 1997; 103:356–62. [PubMed: 9305282]
- Hack M, Flannery DJ, Schluchter M, Cartar L, Borawski E, Klein N. Outcomes in young adulthood for very-low-birth-weight infants. *N Engl J Med.* 2002; 346:149–57. [PubMed: 11796848]
- Inder TE, Buckland L, Williams CE, Spencer C, Gunning MI, Darlow BA, et al. Lowered electroencephalographic spectral edge frequency predicts the presence of cerebral white matter injury in premature infants. *Pediatrics.* 2003; 111:27–33. [PubMed: 12509550]
- Jaynes, ET. *Probability Theory: The Logic of Science.* Cambridge University Press; Cambridge: 2003.

- Jaynes, ET. Probability Theory in Science and Engineering, No. 4. Colloquium Lectures in Pure and Applied Science. Socony-Mobil Oil Co.; 1956.
- Jeffreys, H. Theory of Probability. Clarendon; Oxford: 1939.
- Korotchikova I, Connolly S, Ryan CA, Murray DM, Temko A, Greene BR, et al. EEG in the healthy term newborn within 12 hours of birth. *Clin Neurophysiol.* 2009; 120:1046–1053. [PubMed: 19427811]
- Korotchikova I, Stevenson NJ, Walsh BH, Murray DM, Boylan GB. Quantitative EEG analysis in neonatal hypoxic ischaemic encephalopathy. *Clin Neurophysiol.* 2011; 122:1671–1678. [PubMed: 21334256]
- Liu A, Hahn JS, Heldt GP, Coen RW. Detection of neonatal seizures through computerized EEG analysis. *Electroencephalogr Clin Neurophysiol.* 1992; 82:30–7.
- Löfhede J, Thordstein M, Löfgren N, Flisberg A, Rosa-Zurera M, Kjellmer I, et al. Automatic classification of background EEG activity in healthy and sick neonates. *J Neural Eng.* 2010; 7:016007.
- Mizrahi, EM.; Hrachovy, RA.; Kellaway, P. Atlas of Neonatal Electroencephalography. 3rd ed.. Lippincott Williams and Wilkins; Philadelphia: 2004.
- Navakatikyan MA, Colditz PB, Burke CJ, Inder TE, Richmond J, Williams CE. Seizure detection algorithm for neonates based on wave-sequence analysis. *Clin Neurophysiol.* 2006; 117:1190–1203. [PubMed: 16621690]
- Niemarkt HJ, Andriessen P, Pasman J, Vles JS, Zimmermann LJ, Oetomo SB. Analyzing EEG maturation in preterm infants: the value of a quantitative approach. *J Neonatal-Perinatal Med.* 2008; 1:131–44.
- Niemarkt HJ, Andriessen P, Peters CHL, Pasman JW, Zimmermann LJ, Oetomo SB. Quantitative analysis of maturational changes in EEG background activity in very preterm infants with a normal neurodevelopment at 1 year of age. *Early Hum Dev.* 2010; 86:219–224. [PubMed: 20382486]
- Niemarkt HJ, Andriessen P, Peters CHL, Pasman JW, Blanco CE, Zimmermann LJ, et al. Quantitative Analysis of Amplitude-Integrated Electroencephalogram Patterns in Stable Preterm Infants, with Normal Neurological Development at One Year. *Neonatology.* 2010; 97:175–182. [PubMed: 19864923]
- Palma K, Wikström S, Hippeläinen E, Boylan G, Hellström-Westas L, Vanhatalo S. Detection of 'EEG bursts' in the early preterm EEG: Visual vs. automated detection. *Clin Neurophysiol.* 2010; 121:1015–1022. [PubMed: 20395172]
- Palma K, Stevenson N, Wikström S, Hellström-Westas L, Vanhatalo S, Palva JM. Optimization of an NLEO-based algorithm for automated detection of spontaneous activity transients in early preterm EEG. *Physiol Meas.* 2010; 31:N85–N93. [PubMed: 20938065]
- Roessgen M, Zoubir AM, Boashas B. Seizure detection of newborn EEG using a model-based approach. *IEE Trans Biomed Eng.* 1998; 45:673–85.
- Scher MS, Steppe DA, Banks DL, Guthrie RD, Scwabassi RJ. Maturational trends of EEG-sleep measures in the healthy preterm neonate. *Pediatr Neurol.* 1995; 12:314–322. [PubMed: 7546004]
- Sivia, DS.; Skilling, J. Data Analysis: A Bayesian Tutorial. 2nd ed.. Oxford University Press; Oxford: 2006.
- Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with Support Vector Machines. *Clin Neurophysiol.* 2011; 122:46, 4–473.
- Victor S, Appleton RE, Beirne M, Marson AG, Weindling AM. Spectral analysis of electroencephalography in premature newborn infants: normal ranges. *Pediatr Res.* 2005; 57:336–41. [PubMed: 15635050]
- West CR, Harding JE, Williams CE, Gunning MI, Battin MR. Quantitative electroencephalographic patterns in normal preterm infants over the first week after birth. *Early Hum Dev.* 2006; 82:43–51. [PubMed: 16169163]
- Wilson SB. A neural network method for automatic and incremental learning applied to patient-dependent seizure detection. *Clin Neurophysiol.* 2005; 116:1785–95. [PubMed: 16005680]

- Wood NS, Marlow N, Costeloe K, Gibson AT, Wilkinson AR. Neurologic and developmental disability after extremely preterm birth. EPICure Study Group. *N Engl J Med*. 2000; 343:378–84. [PubMed: 10933736]
1. Bretthorst, GL. *Lecture Notes in Statistics*. Springer-Verlag; Berlin: 1988.
 2. Sivia, DS.; Skilling, J. *Data Analysis: A Bayesian Tutorial*. 2nd ed.. Oxford University Press; Oxford: 2006.

Highlights

1. Developmental features of interest in the EEG recordings from prematurely-born infants can be detected automatically.
2. The automated detection of certain features is as consistent as the inter-rater reliability of experienced electroencephalographers.
3. The method of detection, which uses Bayesian probability theory, can be generalized to identify additional EEG features of interest.

(a) Delta brushes (DB) and smooth delta waves (SDW)

(b) Interburst interval with beginning and end marked with arrows

Figure 1.

Examples of delta brushes, smooth delta waves, and interburst intervals

Figure 2.

Examples of patterns from different classes. In each subfigure, the solid line indicates the most probable value of the amplitude and the dashed line indicates the bounds on the amplitude. Figure 2(a) contains the pattern for the interburst interval class. Figure 2(b) contains the pattern for the general activity class. Figure 2(c) contains one example pattern from the delta wave class.

Figure 3.

Examples of patterns from the delta wave class. In each subfigure, the solid line indicates the most probable value of the amplitude and the dashed line indicates the bounds on the amplitude. The patterns differ in their width and vertical offset to model the various morphologies of delta waves.

Table 1

Comparison of the total number of delta waves and interburst intervals identified by both readers and the algorithm. The algorithm, designed to be conservative in its classification, identified fewer delta waves and interburst intervals than either reader.

	Delta Waves	Interburst Intervals
Reader 1	709	65
Reader 2	754	62
Algorithm	686	59

Table 2

Comparison of the performance of the algorithm to Reader 1 and Reader 2 when identifying delta waves. The algorithm had a lower true and false positive rate when compared to either reader.

(a) Reader 1 as gold standard		
	True Positives	False Positives
Algorithm	557/709 (79%)	129/686 (19%)
Reader 2	579/709 (82%)	174/754 (23%)

(b) Reader 2 as gold standard		
	True Positives	False Positives
Algorithm	574/754 (76%)	112/686 (16%)
Reader 1	579/754 (77%)	129/709 (18%)

Table 3

Comparison of the performance of the algorithm to Reader 1 and Reader 2 when identifying interburst intervals. The algorithm had a lower true positive and false positive rate when compared to either reader.

(a) Reader 1 as gold standard		
	True Positives	False Positives
Algorithm	47/65 (72%)	12/59 (20%)
Reader 2	48/65 (74%)	14/62 (23%)

(b) Reader 2 as gold standard		
	True Positives	False Positives
Algorithm	47/62 (76%)	12/59 (20%)
Reader 1	48/62 (77%)	17/65 (26%)

Table 4

Comparison of the performance of the algorithm to Reader 1 and Reader 2 when distinguishing between smooth delta waves and delta brushes. The algorithm did not perform as well as the readers, with smooth delta wave and delta brush true positive rates lower than both readers.

(a) Reader 1 as gold standard		
	SDW True Pos.	DB True Pos.
Algorithm	375/432 (87%)	35/66 (53%)
Reader 2	386/432 (89%)	44/66 (67%)

(b) Reader 2 as gold standard		
	SDW True Pos.	DB True Pos.
Algorithm	358/408 (88%)	42/90 (47%)
Reader 1	386/408 (95%)	44/90 (49%)

Table 5

The performance of the algorithm when Reader 1 and Reader 2 are in agreement.

	True Positives	False Positives
Interburst Intervals	42/48 (88%)	7/59 (12%)
Delta Waves	498/579 (86%)	53/686 (8%)
Smooth Delta Waves	344/386 (89%)	17/406 (4%)
Delta Brashes	27/44 (61%)	42/92 (46%)