2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782611

Spike isolation from background signal in neonatal EEG data using an integrated independent component analysis method

Mina Ezzati Asl*, Josue D. Rodriguez †, Mercedes Cabrerizo‡ ,Anuj Jayakar§
Armando Barreto¶, Malek Adjouadi‖
*†‡¶‖Department of Electrical and Computer Enginnering, Florida International University, Miami, Florida
§ Nicklaus Children's Hospital, Miami, Florida
Email: {*mezza004,†jrodr1309,‡cabrerizo,¶ barretoa,‖ajouadi}@fiu.edu, §anuj.jayakar@gmail.com

***Abstract*—Spike detection in epileptic neonates is a challenging task since the recorded electroencephalography (EEG) data are often fraught with artifacts and noise. This study aims to enhance the clarity of epileptic spikes by separating them from background activity using an integrated method based on the independent component analysis (ICA). We analyzed spikes of 12 epileptic neonates as marked in their EEG scalp recordings by a clinical expert. The proposed method makes use of the ICA method to isolate the source of the spikes and then apply a power frequency analysis and template matching to validate the performance of the ICA. Isolating a spike is achieved by choosing the component that should correspond to its defining characteristics, followed by signal reconstruction using that component. To evaluate the accuracy of our spike isolation method, we first check if the power spectrum of the separated spikes aligns with the typical power spectral density observed in neonates. Subsequently, we measured the degree of similarity between the extracted spike and a predefined spike template, comparing it against the original spike segment. With this integrated method, the results show the successful extraction of 29 out of the 37 marked spikes (i.e., 79 percent), which signifies that ICA can serve as a promising approach in the initial isolation process of spikes in EEG records of neonates. This could lead to further investigation into those subtle features or changes missed on those EEG records of the marked spikes that were not separated. Determining such features and subtle changes, if indeed inherent to spikes, could lead to the development of enhanced spike detection methods in neonates. It should be noted that in 5 out of the 37 epochs, we could not identify any independent component as a spike source, and 3 out of 32 remaining cases showed unsuccessful separation in validation, possibly due to the source not being statistically independent or being Gaussian in nature. In such cases, the expert clinician(s) could review or reconsider marking such spikes.**

***Index Terms*—Independent component analysis, Spikes, Neonatal seizure, Power spectrum analysis, template matching, Dynamic time wrapping**

I. Introduction

Neonatal seizures are a critical neurological condition in newborns. Continuous multi-channel EEG monitoring remains the only reliable method for detecting all neonatal seizures [1], including both ictal (seizure) events and interictal (spike) activities. However, interpreting EEG in neonates is challenging due to the difficulty in differentiating between normal brain activity and pathological conditions in the immature neonatal brain [2]. This complexity requires a precise and specialized approach to ensure accurate diagnosis and effective monitoring.

The accurate detection and analysis of spikes in neonatal EEG recordings are also critical for enhancing seizure prediction and management. Misidentification of these spikes can lead to either a high rate of false alarms or a low seizure detection rate [3]. To address this issue, our study focuses on the application of Independent Component Analysis (ICA) as a method for isolating spikes from background signals and the ubiquitous noise in neonatal EEG recordings. We aim to improve spike detection and gain a deeper understanding of spike characteristics in neonates. This could potentially lead to better seizure prediction and more effective management of neurological conditions in newborns.

In our study on spike separation using Independent Component Analysis (ICA), we explore the potential of ICA as a novel technique for isolating epileptic discharges from scalp EEG recordings of neonates. Epileptic discharges are often entangled with background brain activity and artifacts, posing a challenge for accurate detection and analysis. ICA, by design, separates statistically independent components from mixed data. Given that interictal epileptic discharges in EEG are generally infrequent and independent from background activity [4], we propose that ICA is an ideal method for their isolation. Through visual analysis of the separated components, we aim to distinguish epileptic spikes from background activity [5].

To validate the effectiveness of our spike isolation method in removing noise and artifacts, we used the characteristics of power spectral density. The power spectrum of neonatal EEG is distinctively characterized by a dominant frequency within the delta and theta frequency bands. When analyzing the neonatal EEG, the full spectrum is considered, based on the assumption that it follows an inverse power law [7]. This is highlighted by the fact that the majority of spectral energy in newborn EEG is concentrated in the lower frequency bands [7]. If the isolation process is successful, the power spectrum of the isolated spikes should be consistent with the expected spectral distribution in neonates.

For the second validation of the separated spikes, the similarity between a spike template and the isolated spike is measured using the dynamic time wrapping method and then compared to the similarity between the template and the whole spike segment. If the isolation were performed effectively, the similarity between the extracted spike should be considerably higher when compared with the entire spike epoch.

II. Methodology

A. Data

The EEG data was acquired at Nicklaus Children's Hospital, Miami, Florida, using the 10-20 system. The electrodes considered are C3, C4, O1, O2, CZ, F3, F4, F7, F8, FZ, FP1, FP2, FPZ, P3, P4, Pz, T3, T4, T5, T6, and the ground electrode was FPz, and the sampling frequency is 256 Hz. The power noise line is removed from the EEG signal using a bandpass filter with cut-off frequencies of 0.5 and 40 Hz. This filtering range is based on the standard of newborn EEG analysis [7], as the spectral areas of interest in neonates in the power spectrum are delta (0.5-4Hz), theta (4-8Hz), alpha (8-13Hz), Beta (13-30Hz) and gamma (30-40Hz). The data consisted of EEG scalp recordings of 12 newborn patients, in which the spikes were marked by a neurologist. A total of 37 epochs (3 sec) containing spikes were extracted from all patients for implementing and testing the proposed spike isolation method. The spatial distribution of all epochs is baseline-corrected by subtracting the mean value which has a similar outcome as using the average reference [6]. This helps in isolating true brain activity from common noise or bias present in the recorded EEG signals.

B. Independent Component Analysis

The EEG scalp recording is a linear mixture of electrical activity from various parts of the brain. Given that epileptic discharges are likely independent of background activity, we employ ICA to decompose EEG segments into independent components [4]. It is worth noting that before applying ICA to the EEG signal, we do not have access to its sources to inspect if there are any Gaussian sources or to test if they are statistically independent. So, among the various implementations of ICA, we chose an algorithm based on self-organizing learning, the Infomax algorithm, which operates without prior knowledge of the signal sources [8].

It is important to note that ICA is effective in separating non-Gaussian sources, so it cannot distinguish perfectly Gaussian sources. Even when the sources are not independent, ICA finds a space where they are maximally independent. This algorithm is implemented in *runICA* function which we used via EEGlab 2023.0 toolbox in MATLAB. To have an insight into how the ICA algorithm works on multichannel data, consider that our data is X, and components are a set of vectors that project the original data (X) onto new axes found by ICA. The weight matrix W transforms data from the original space to the source space. So ICA algorithm tries to find W that will give Y which is the best estimation of S. With S representing the source activity, the relationship is given by:

$$Y = W.X \approx S \tag{1}$$

The Infomax algorithm computes W, the unmixing matrix, using the following steps [9]:
1. Initialize $W(0)$ a random value.
2.
$$W(t+1) = W(t)\eta(t)(I - f(Y)Y^T)W(t) \tag{2}$$

Where $\eta(t)$ is a function that specifies the step size and $F(Y)$ is a general function chosen based on the type of distribution.
3. If convergence has not been achieved, return to the second step [9].
Once the W matrix is computed, we have our source components. However, the activity of the brain source is unitless unless it is projected onto the electrodes. Each source creates a contribution at each electrode site, to reverse the projection of one component to the electrode space we use W^{-1} as the inverse of the weight matrix to go from the source space S to the data space X as follows:

$$X = W^{-1}.S \tag{3}$$

C. Spike separation from background signal

The ICA algorithm is deployed over all 37 spike segments marked by the expert clinician. Since we have 19 channels, ICA gives us 19 components and each one potentially represents a unique source of the recorded EEG. These sources could come from neural and artifact sources. We could recognize the epileptic discharge in the decomposed components by visual inspection [6]. A component corresponding to the spike source should replicate the spike's shape at the exact same time when the spike happens in the EEG segment. In other words, their peaks in the time domain should be aligned. "Fig. 1" illustrates a sample of the spike epoch, and "Fig. 2" displays the components of that segment in which the spike component is highlighted.

This step requires careful observation of the spike segment and its components. For all 37 spike segments, components and spike segments have been compared in the same time span to examine if any components match the spike. However, in some cases, it might not be immediately evident if a component matches the spike's shape, or the spike's form may not be distinctly observable in the EEG segment. So in these cases, first we removed some obvious non-spike components and then reconstructed the signal with the remaining components. Then, with a clearer version of the spike segment and the remaining components, we do a visual examination to find the spike's component.

Typically, we can ascertain if one of these components occurs simultaneously with the spike and have a similar shape, such a component can be considered a candidate for the spike source. After selecting the spike source among various alternatives, including background signals and noise, we use that single component to reconstruct the segment. This process results in a signal comprising solely of the spike source. The reconstructed segment, which could potentially represent the isolated spike, is shown in “Fig. 3”. The next step involves analyzing the effectiveness of this method in isolating spikes.

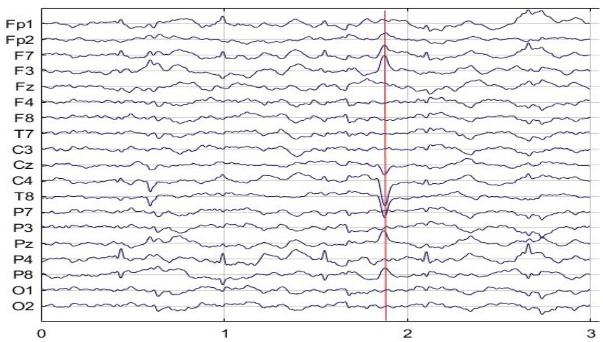


Fig. 1: A sample of a spike segment where the spike is pointed in the figure.

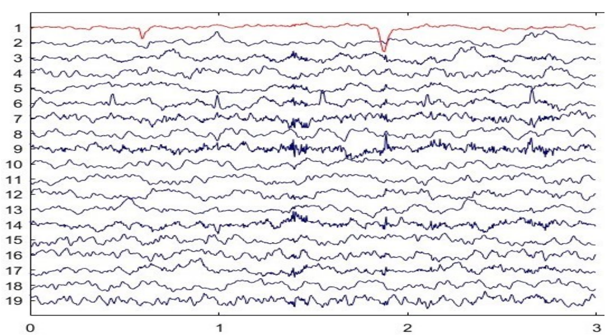


Fig. 2: All the components that are obtained by applying the ICA algorithm on the spike segment

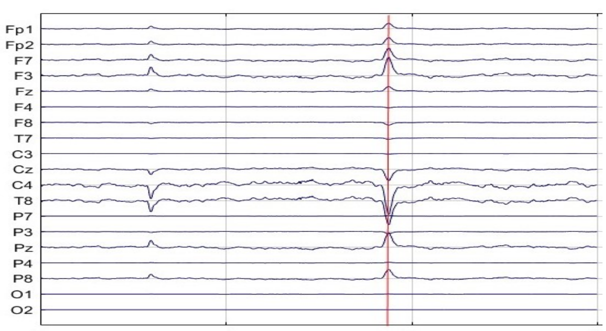


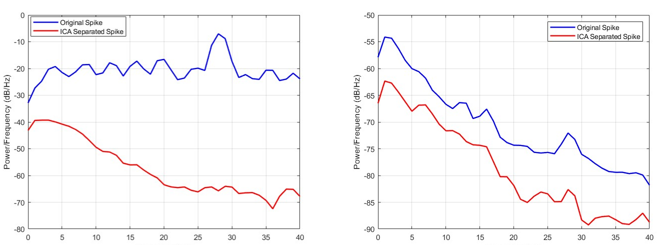
Fig. 3: Separated spike by reconstructing segment using the spike’s source component

D. Validation

To evaluate the precision of the chosen component, and to investigate if the spike is separable by ICA, we consider two methods to analyze the isolated spikes. The first method involves analyzing the power spectrum, assuming the spectral power of neonatal EEG to be usually concentrated in lower frequency bands such as delta and theta. It follows the inverse power law, which means that as the frequency increases, the power decreases [7]. We use this characteristic of the power spectrum as a criterion for the successful isolation of spikes.

To compute the power in all frequency bands, considering the sampling frequency is 256 Hz, we employ Welch’s method via MATLAB’s *Pwelsh* function to calculate the power spectral density (PSD) of the signal. Then, the power within each frequency band is computed using the band-power function, which integrates the PSD over the specified frequency range. By comparing the spectral power of both the original and reconstructed spike epochs, we observed that in some patients with high amplitude and high-frequency signals in their EEG recording, the power in beta and gamma range is as high as delta and even higher, which is not consistent with the neonate expected spectral density.

Successful isolation of the spike source should result in a spectrum dominated by lower frequencies, particularly in the delta range, indicating correct separation from background signals and noise. “Fig. 4a”. Also, in the cases with normal spectral power, spike separation results in a more distinct reduction in high frequency rather than low frequency “Fig. 4b”. Power distribution also varies comparing spike source and other components, for spikes power spikes tended to be localized to one or a few electrodes, whereas power for other components was more distributed “Fig. 5”.



(a) noisy spike segment

(b) normal spike segment

Fig. 4: Comparison of power spectrum between original spike segment and the isolated spike of 2 cases. The first sample shown in part (a) represents a spike segment containing high-frequency and high-amplitude signals, resulting in higher spectral power in the beta and gamma bands. The second sample shown in part (b) corresponds to a normal spike segment. In both cases, spike isolation leads to a reduction in power at higher frequencies, and the spike source exhibits descending power as the frequency increases.

In the second step, we involved template matching to evaluate the clarity of the spike isolated by ICA. To create a general template, we used a data set of 103 epileptic patients obtained from Baptist Hospital, Miami, Florida, from which the spike segments had already been marked and extracted by a clinical expert. After aligning and averaging these spikes, we formed our template. Now, for measuring the similarity between this template with the original spike segment and the reconstructed spike segment, we use dynamic time warping (DTW). The DTW method is often used to calculate the similarity between two sequences, which may vary in length or speed. The process of DTW involves mapping points from one sequence to another such that the sum of the distances between these matched points (usually Euclidean distances) is minimized. A higher similarity between the isolated spike segment and the spike template, as compared to the original spike segment and the template, would indicate correct component selection and successful spike isolation. For illustrative purposes, “Fig. 6” shows how the template and the spike segment are overlayed on each other.

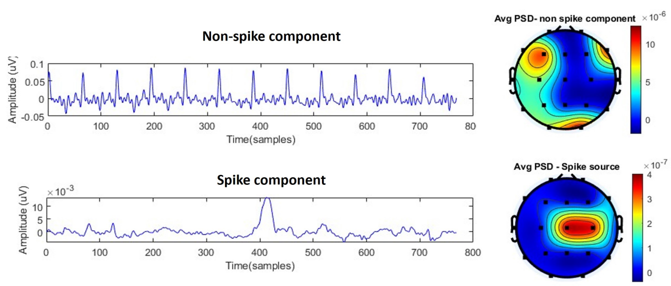


Fig. 5: In this figure, the power distribution of a reconstructed segment with the spike’s source component and a reconstructed segment with a non-spike source component are compared. The spike’s component power is centered in one or a few electrodes, but for the non-spike component, the power is spread over multiple electrodes.

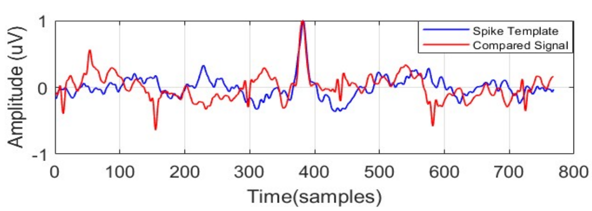


Fig. 6: Spike template and a spike segment overlaid on each other to illustrate the similarity measurement using DTW method.

III. Results and Discussion

It is important to acknowledge that not all spikes can be separated using this method; only those spike sources that are statistically independent can be effectively isolated from other signal sources. This is why the component must closely resemble the exact shape of the spike. In some cases, a spike may result from the aggregation of multiple sources, not just one. After applying ICA to all spike epochs, power frequency analysis and similarity measurements were conducted on the isolated spikes. Out of the 37 spike segments analyzed 29 spikes appeared to be separable by a single component.

In approximately 5 out of the 37 epochs, we could not identify any component as a spike source, possibly due to the source not being statistically independent or being Gaussian in nature. For the 32 spikes that were isolated, spectrum analysis was performed, and appeared consistent with the inverse power law, showing dominance in the delta frequency range. For the second validation of 32 spikes, the similarity between this template and both the reconstructed spike and the original spike segment has been measured using the Dynamic Time Wrapping method (DTW). The DTW hence enables the measurement of pattern similarity in two signals. A lower distance indicates higher similarity. The Distance measurement is shown in “Fig. 7”.

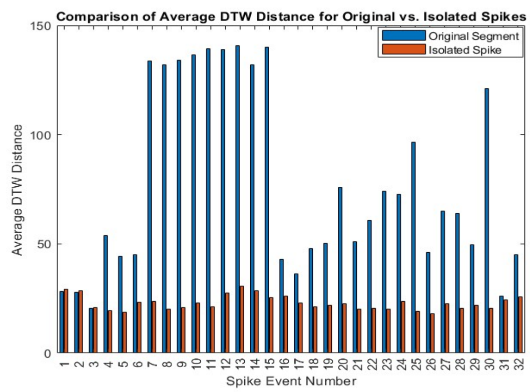


Fig. 7: Distances between the template and original spike segment vs distance between the template isolated spikes for all 32 spikes. We can observe that the first 3 isolated spikes, despite other spikes, have higher distances with the template compared to their original format, indicating unsuccessful spike separation for these 3 cases. However significant distance reduction was achieved by spike separation for other cases, which means better similarity with the template and successful spike isolation.

As observed from the chart, the noisy segments exhibit limited similarity to the template due to the presence of high-frequency signals, resulting in a significant distance between the template and the spike segment. However, upon separating the spike’s source, the distance decreases to a range of [18.0412 - 30.6194], with an average of 22.9140 $\pm$ 3.3188, indicating a notable increase in similarity between the isolated spikes and the template in most cases. This suggests that the selected component effectively isolates the spike patterns from other signals within the EEG segment. Out of 32 spikes, for 3 spikes, we can see spike separation was not effective even though the original segment has more similarity with the template. So by excluding those 3 spikes, for 29 spikes, approximately on average this method increases the similarity 3 times, which indicates how this extraction enhances spike analysis in future studies.

IV. Conclusion

We investigate the capability of independent component analysis in isolating spikes from background signals in EEG recordings of epileptic neonates. We analyzed 37 spikes of 12 patients, 29 of them appeared separable using independent component analysis. Power spectrum analysis revealed that the separated spikes predominantly exhibited power in the lower frequency ranges, which is consistent with the expected power distribution in neonates. This suggests that the power associated with high-frequency artifacts and noise was effectively eliminated from the signal. Furthermore, to have secondary evidence of correct isolation of the spike, the similarity between a spike's template and the extracted spike is compared with the similarity between the template and the EEG segment containing the spike by the Dynamic Time Wrapping method. The results show that the distance between the template and the isolated spike was significantly less than the difference between the template and the original EEG segment, which implies that ICA could be an effective way to extract spikes from background signals or any inherent noises. This method is particularly valuable in the context of neonatal EEG, which is fraught with artifacts and noise, making spike detection challenging. As this method improves the quality of spikes that are detected, it will enhance the precision of feature extraction of spikes and as a result, it will improve spike detection methods and help seizure prediction algorithms in the future. The focus could be placed on those features deemed relevant, and to search for other subtle features that were missed with those EEG segments with the few spikes that were not separated successfully.

V. Acknowledgement

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-2018611, and CNS-1551221. We also acknowledge the support of the Ware Foundation.

References

[1] Temko, A., Marnane, W., Boylan, G. and Lightbody, G., 2015. Clinical implementation of a neonatal seizure detection algorithm. Decision support systems, 70, pp.86-96.

[2] Nayak CS, Anilkumar AC. Neonatal EEG. [Updated 2023 Apr 16]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536953/

[3] Hassanpour, H. and Mesbah, M., 2003, July. Neonatal EEG seizure detection using spike signatures in the time-frequency domain. In Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings. (Vol. 2, pp. 41-44). IEEE.

[4] Kobayashi, K., James, C.J., Nakahori, T., Akiyama, T. and Gotman, J., 1999. Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. Clinical neurophysiology, 110(10), pp.1755-1763.

[5] Urrestarazu, E., Iriarte, J., Artieda, J., Alegre, M., Valencia, M. and Viteri, C., 2006. Independent component analysis separates spikes of different origin in the EEG. Journal of clinical neurophysiology, 23(1), pp.72-78.

[6] Kobayashi, K., Merlet, I. and Gotman, J., 2001. Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. Clinical Neurophysiology, 112(3), pp.405-413.

[7] Rankine, L., Stevenson, N., Mesbah, M. and Boashash, B., 2006. A nonstationary model of newborn EEG. IEEE Transactions on biomedical engineering, 54(1), pp.19-28.

[8] Bell, A.J. and Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), pp.1129-1159.

[9] Langlois, D., Chartier, S. and Gosselin, D., 2010. An introduction to independent component analysis: InfoMax and FastICA algorithms. Tutorials in Quantitative Methods for Psychology, 6(1), pp.31-38.