Spike sorting using Superlets: Identifying feature importance through perturbation

Eugen-Richard Ardelean
Computer Science
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
ardeleaneugenrichard@gmail.com

Răzvan-Dan Terec
Computer Science
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
razvan.terec@yahoo.com

Codruța-Maria Marieș
Computer Science
Technical University of Cluj-Napoca
Cluj-Napoca, Romana
codruta.maries24@gmail.com

Vasile Vlad Moca
Transylvanian Institute of Neuroscience
Cluj-Napoca, Romania
moca@tins.ro

Raul Cristian Mureșan
Transylvanian Institute of Neuroscience
STAR-UBB Institute
Cluj-Napoca, Romania
muresan@tins.ro

Mihaela Dînșoreanu
Computer Science
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Mihaela.Dinsoreanu@cs.utcluj.ro

***Abstract*— Spike sorting is a technique in the field of neuroscience applied to identify and classify the action potentials (spikes) of neurons in extracellular recordings. This method is important considering that the simultaneous activity of several neurons can be recorded through the use of electrodes, but the recorded signals are often mixed. Extracting the characteristics of individual spikes provides valuable information about the identity of neurons that generate them, the type of cells involved (inhibitory/excitatory), encoding of information in spike rates or spike times, and so on. Therefore, unmixing the spikes recorded on an electrode via spike sorting is crucial. Here, we apply a novel method to extract features of mixed spikes, namely the Superlet transform, which enables the computation of spectral characteristics with a higher resolution. We use machine learning to determine which features of the Superlet spectrum contain the most information about the shapes of individual spikes, thereby enabling their unmixing during spike sorting.**

Keywords—spike sorting, superlet, bicubic interpolation, disturbance of characteristics, performance metrics, neural network

I. INTRODUCTION

Spike sorting [1] is the process of identifying and classifying neuronal discharges, also called spikes, recorded as electrical signals, aiming to assign each spike to a specific source neuron. This technique provides important information that facilitates the subsequent analysis of neuronal activity. By following the spikes recorded over time, one can observe details about the response of neurons to stimuli, the coding of information, the synchronization of neuronal activity, but also other important neurophysiological processes. Furthermore, the sorting of spikes allows the collection and analysis of the data necessary for constructing and validating neural network models. These models help to understand and to predict the behaviour of neural networks in various experimental conditions.

Previous work [2] has demonstrated that the Superlet Transform is an improvement compared to state-of-the-art feature extraction methods in terms of the separability of clusters in the spike sorting process. Therefore, here we aim to identify which are the most important features of the spectrogram, obtained by applying the Superlet Transform, that enable good separability of different clusters of spikes.

II. THEORETICAL FOUNDATIONS

A. Challenges

Neuronal spikes are the main way neurons transmit and encode information in the nervous system. They are used to communicate with other neurons and transmit signals between different brain areas. Therefore, the analysis and recording of neuronal spikes are essential methods in neuroscience to understand neuronal behaviour and brain function.

In extracellularly recorded neuronal data, a cluster consists of several spikes having similar characteristics and coming from the same neuronal source. The structure of the clustering space depends critically on the features that are extracted from spikes and used for the clustering. Most feature extraction algorithms do not ensure perfect separability between data classes. In general, feature extraction algorithms are designed to shift the raw data into a new feature space, even a reduced space, while preserving as much information as possible, but there is no guarantee that these features will offer perfect separability. Class separability is an important factor in the classification and the clustering of data after feature extraction. Even if the extracted features are not perfectly separable, a powerful classification algorithm might still be able to distinguish between classes and achieve good performance.

The Superlet transform [3] has shown that it can provide improved cluster separability in the spike sorting process compared to classical feature extraction methods [2]. The Superlet transform provides a frequency resolution adapted to the signal characteristics and allows precise time-frequency localization of spikes, which means that it can better identify and separate the spikes in the time-frequency space.

To understand why the Superlet transform enables better feature separability, we used real datasets recorded from the brain of adult mice under anaesthesia through 'in vivo' electrophysiology. Datasets contain the activity of neurons in response to different stimuli captured with electrodes implanted in visual cortex. For the datasets used in this study, 32-linear probes (Cambridge NeuroTech) were employed, and neuronal activity was recorded at 32 kSamples/s (Multi Channel Systems MCS GmbH). Each electrode has a slightly different location once inserted into the brain, thus each channel contains a varying number of spikes as each has

979-8-3503-7035-5/23/$31.00 ©2023 IEEE

captured the activity of different neurons from its vicinity. Furthermore, the data has been manually spike sorted by an expert in order to provide a 'ground truth' through which classification and comparisons can be made.

B. *Superlet*

The Superlet transform [3], initially developed for the analysis of brain signals, relies on multiple Wavelet transforms to compute a representation of signals in the time-frequency domain.

In the process of spike sorting [1], the Superlet transform has been used in the stages of extracting the characteristics of the action potentials, or spikes [2]. After applying the transform, several characteristics can be extracted, such as the time-frequency span and magnitude or power.

The formula of the Superlet Transform [4] is:

$$SLT_{X,O,c_1}(t,f) = \left[\prod_{i=1}^{o} CWT\,(t,f,c_1)\right]^{\frac{1}{o}} \quad (1)$$

where x represents the signal, o represents the order of the Superlet, c_1 is the number of cycles of the base Wavelet [4], t represents time and f represents frequency. The *CWT* stands for Continuous Wavelet Transform.

Therefore, the Superlet method is a spectral estimator useful for signal processing, which combines characteristics of short and wide wavelets to isolate processes in the time-frequency domain.

C. *Bicubic interpolation*

Bicubic interpolation [5] is used to estimate values between neighbouring data points in a two-dimensional representation, using a cubic function. Thus, the time-frequency representations resampled using the bicubic interpolation technique will have a uniform and continuous resolution. This resampling can be beneficial for certain applications, such as feature extraction. It also facilitates data alignment in a common format to perform quantitative and qualitative analyses on the recorded signals.

For a coordinate matrix (N, 0) x (0, N), the formula for its interpolated surface is:

$$\mathrm{P}(\mathrm{x},\mathrm{y}) = \sum_{i=0}^{N}\sum_{i=0}^{N} a_{ij}\, x^i y^j \quad (2)$$

where the parameters x and y are the coordinates of the point of interest, while a_{ij} is the applied interpolation factor [6].

D. *Feature permutation*

Feature permutation is a technique that can be used to evaluate the importance of features in classification. This involves disturbing certain features of the data that will be used to train a classifier and evaluating the performance of the classifier [4] with the help of performance metrics, such as Accuracy, or the F1 Score. The drop in performance when a feature is perturbed will indicate the amount of information that the classifier is able to learn through that feature. The permutation process involves reorganizing all the values of a characteristic of the data across all samples, such that no value corresponds to its own sample. If the feature is important for classification, the permutation is expected to lead to the performance degradation of the classifier. A limitation of this method is marked by the fact that the disturbance of a feature may not affect the performance of the classifier if other features are strongly correlated with it, because a compensation phenomenon occurs.

Thus, the permutation of features allows the identification of important features for classification and should generally lead to a decrease in performance, but there may be cases where performance increases or remains largely unaffected. To tackle this issue, we compute the sets of correlated features and perturb them together, to avoid performance compensation (see below).

III. SOLUTION

In carrying out the proposed method of identifying the most important features areas for the learning of the model through the perturbation of characteristics, the steps of the proposed method are shown in Figure 1.

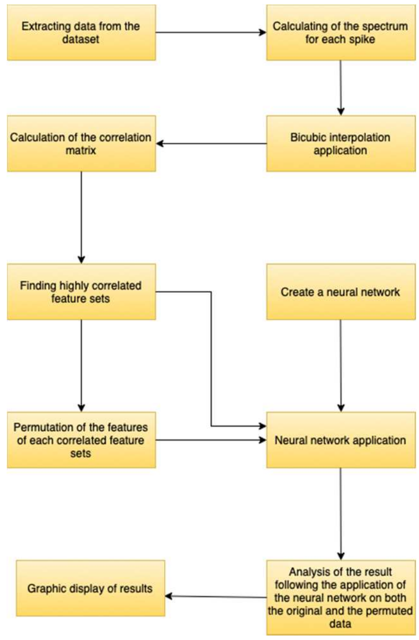

Figure 1. The flow of the correlated feature set perturbation method.

The first step of the proposed pipeline is the computation of the Superlet spectrum for each spike. Downsampling of the spectra, through a bicubic interpolation, is then applied to reduce the dimensionality and to reduce the computation time, but the steps of the pipeline are viable even without the bicubic interpolation. As the perturbation of a single feature may not be enough due to the compensation provided by correlated features in the training process, we chose to instead perturb all correlated features together. Therefore, the next step of the pipeline is to find the sets of features that are correlated above a certain threshold. This is accomplished through the processing of the correlation matrix. Once the correlated feature sets have been extracted the comparison can commence. An instance of the neural network is first trained on the unperturbed features. Next, each individual set of correlated features is perturbed, and a neural network of the same architecture is trained on the modified dataset that contains the perturbation of a single feature set. Through the difference in performance of the neural networks trained on the unperturbed and perturbed dataset, the impact in learning of the feature set is obtained. The following subsections provide a detailed description of these steps.

A. *The time-frequency spectrum*

As the input of the classifier (neural network model) is the spectrogram obtained by applying the Superlet Transform, it is the characteristics of the spectrogram that will be perturbed.

Figure 2 shows the time-frequency spectrum of three average spike waveforms using the Superlet method (with order $o=2$, and the number of cycles $c_1=1.5$) on a set of real data. At the top of each figure, the average spike from each cluster can be seen measured in mV, and at the bottom we show the representation in the time-frequency space of the spike. The classical frequency interval of spikes is the 300-3000Hz range [1], we have chosen to use the 300-7000Hz range as the original range would cut off the spectral representation of the spike.

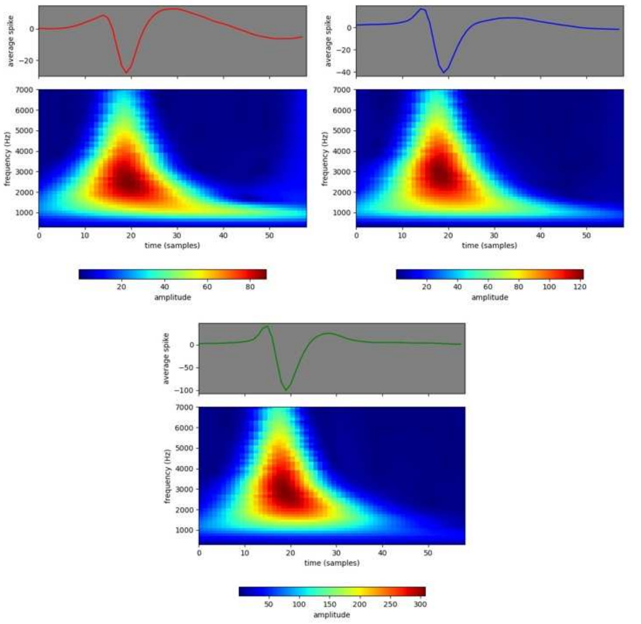

Figure 2. Average spikes and their corresponding spectrograms for 3 different clusters (neurons) of a single channel of real data.

B. *Bicubic interpolation and the correlation matrix*

Our initial step was to resize the spectrogram matrix from the initial size of [58, 50] to a size of [14, 12]; thus, obtaining a total of 168 features. This process of rescale is achieved through bicubic interpolation as shown by Figure 3.

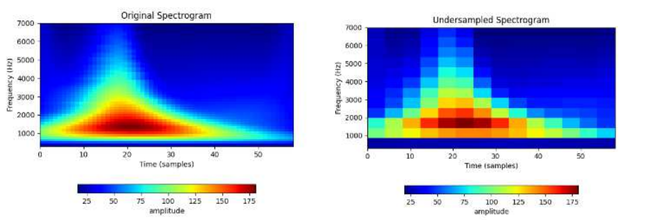

Figure 3. Example of bicubic interpolation for downsampling the spectrograms of spikes.

To graphically visualize how strongly the characteristics of the action potentials are correlated, we calculated the correlation matrix, which shows how all possible pairs of values are closely related to each other for a certain feature correlation threshold. It is important to specify that the correlation coefficient varies between [-1, 1], where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates the absence of a correlation between the coefficients [7].

Figure 4 shows the correlation matrix, from which we can observe the correlations between features represented visually through a heat map, in which the red colour is intended for very high values, and the blue one for low values.

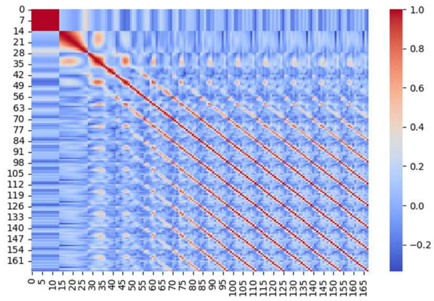

Figure 4. An example of a correlation matrix obtained from real data after the downsampling of spectrograms through bicubic interpolation.

Another step using the bicubic interpolation is taken at the end, where the spectrogram matrix is upscaled from the dimension of [14, 12] to the initial dimension [58, 50], to return to the original dimension. This is shown in Figure 5.

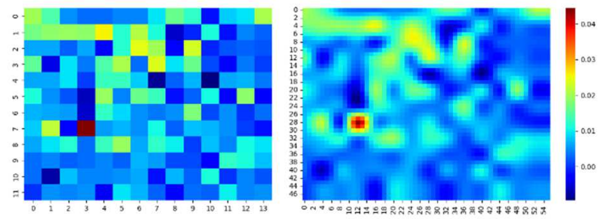

Figure 5. Example of bicubic interpolation for upsampling on the drop in performance for each feature set.

C. *Feature permutation*

One of the most important steps is the feature perturbation method. To analyse how much each feature set matters in learning, we perturbed each feature set in several ways. We chose to disturb the feature sets 10, 20, and respectively 50 times, applying different threshold values, that define what a "correlated set" is. Each feature set contains those features that have a correlation greater than or equal to a certain threshold, and the threshold refers to a value used to decide whether two features are considered correlated or not. After a thorough analysis of the three cases, we noticed significant differences between them. Permuting the features of the spikes 50 times rather than 10 or 20 times, leads to a more stable statistical estimation of the effect of perturbation. Additionally, the more extensive perturbation can help identify features with a greater contribution to correct classification, helping in the elimination of irrelevant features.

To observe the impact of the perturbation method on the characteristics of the spikes, we made a plot bar from which we can deduce the difference in the accuracy metric values between each original and perturbed set. This is shown in Figure 6. Thus, from the analyses made, we noticed that, in most cases, by disturbing the data, the value of accuracy decreases. Evidently, the greater the decrease in accuracy on the original set compared to the disturbed one, the higher the contribution to learning of that set of characteristics.

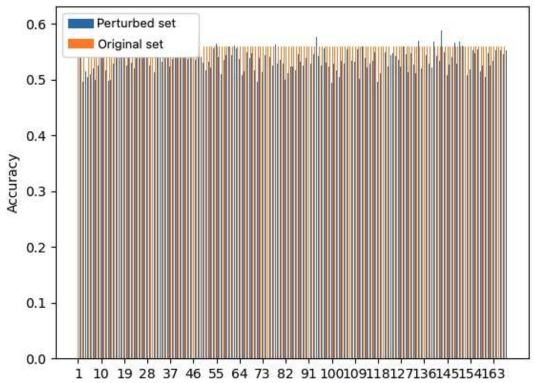

Figure 6. The difference in results between the original set and the perturbed set when applying the Accuracy performance metric for spikes recorded on a single channel in real data (channel 6), using a threshold of 0.5.

D. *Neural network*

To identify the most important characteristics of spikes in automatic learning, they must be used in the training of a classifier. Here, we used a neural network that features an input layer, a hidden layer with the ReLU activation function, and an output layer with the Softmax activation function. The input layer [8] has the role of receiving the input values of each data set, the number of neurons being determined by the size of a sample of the input data. Regarding the output layer [8], it produces the final result of the neural network, both for the original set of features and for the disturbed one, and the activation function returns a vector with the length equal to the number of classes in the [0,1] interval and ensures that the sum of probabilities of all classes is 1. The number of classes for a certain data channel is equal to the number of individual neurons observed on the channel and separated during manual spike sorting in order to obtain the ground truth.

The metrics used to evaluate the performance of the classification model are accuracy and F1 score. Accuracy represents a measure of the correct proportion of predictions out of the total number of predictions, calculated by dividing the number of correct predictions by the total number of test examples. Instead, the F1 Score is a measure of the balance between the precision and recall of a model, being calculated based on these two metrics.

From Figure 7 and Figure 8, in which we applied the accuracy and F1 score performance metrics, we can see that in the first case, we obtain a maximum drop in performance of 3% while in the second case 9%. This implies that, at least in certain situations, the F1 Score is a more robust and informative metric than Accuracy.

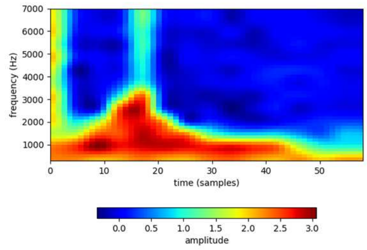

Figure 7. Drop in Accuracy for spikes recorded on a single channel, namely channel 6, from a real dataset, when different time-frequency regions are perturbed. Larger values indicate a larger drop in performance, hence more importance of the respective features.

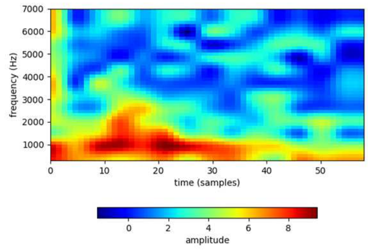

Figure 8. Drop in the F1 score for spikes recorded on channel single channel, namely channel 6, from a real dataset, when different time-frequency regions are perturbed. Larger values indicate a larger drop in performance.

E. *Superlet parameters*

The Superlet Transform has several parameters that may affect performance. Choosing these parameters was done by testing various combinations and comparing results. Thus, we made a comparison between the following parameters: Superlet of order 1 (Wavelet), Superlet of order 2, respectively Superlet of order 5 shown in Figures 9, 10, and 11 respectively on the same real dataset.

Considering these graphical representations, we found that the Wavelet has a good temporal resolution, but the 2nd-order Superlet offers the best precision across multiple areas such as low frequencies, high frequencies, and transient frequencies in time.

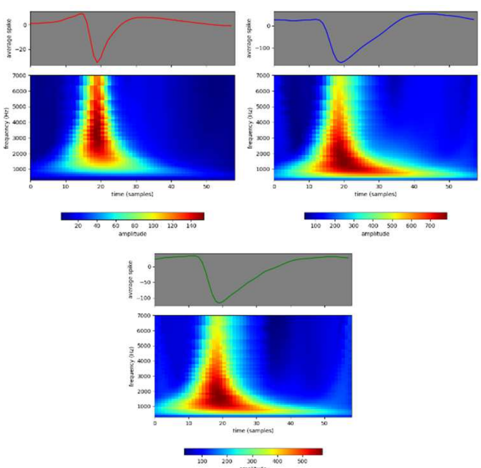

Figure 9. The average spike per cluster and the time-frequency spectrum for Superlet of order 1 (wavelet) and number of cycles 1.5 on real data.

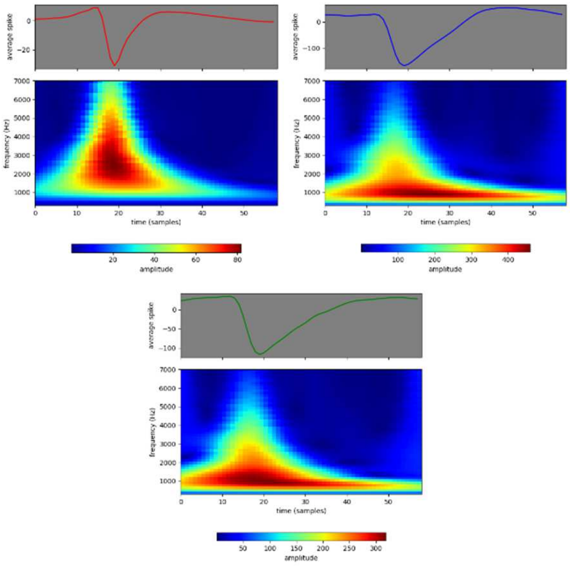

Figure 10. The average spike per cluster and the time-frequency spectrum for Superlet of order 2 and number of cycles 1.5 on real data.

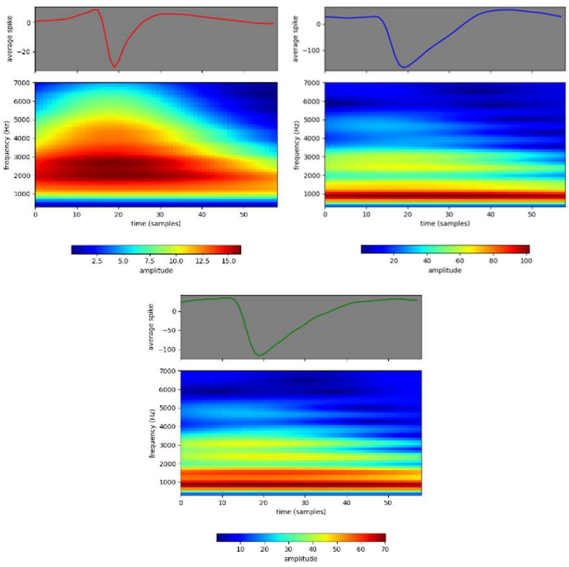

Figure 11. The average spike per cluster and the time-frequency spectrum for Superlet of order 5 and the number of cycles 3 on real data.

From the analyses carried out, it can be easily inferred that the Superlet of order 5, compared to the Wavelet or Superlet of order 2, has a decrease in precision over time, but at the same time, an increase in precision for the frequency domain. In comparison, the Wavelet has a good temporal precision but loses its precision in frequency. Thus, the best precision in both time and frequency is offered by the Superlet of order 2. This is easily visualized in Figure 12 where the drop in performance is shown for the Superlet of order 1 on the top-left, of order 2 on the top-right and of order 5 on the bottom regarding the F1 Score.

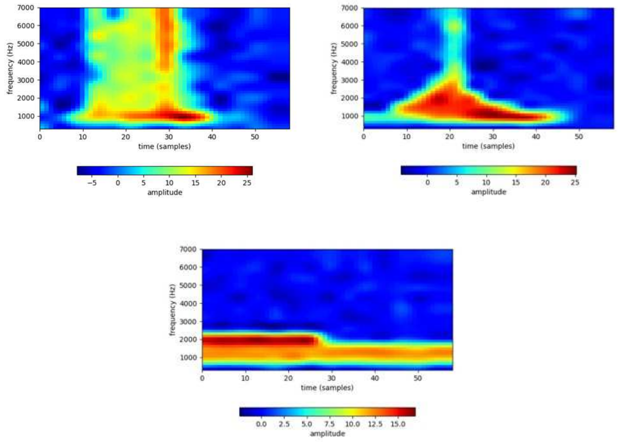

Figure 12. Drop in the F1 Score for spikes recorded on channel 17, for Superlet of order 1 (top-left), of order 2 (top-right) and of order 5 (bottom).

IV. RESULTS

From the obtained graphic representations, we created a series of tables to have a better comparative perspective on the results obtained, according to certain criteria.

A. *Evaluation of the performance of the characteristics depending on the threshold*

From the comparative analysis we performed in Table I, we can note that when we identify the correlation sets, applying a threshold of 0.3, we obtain a higher performance than in the other cases. This is because the lower the threshold value, the larger the correlated feature sets.

Table I. Comparative analysis of the difference in performance when training on the original and perturbed sets depending on the performance metric and correlation threshold applied to 3 channels on real data.

CHANNEL	**ACCURACY**			**F1 SCORE**		
	thr = 0.3	*thr = 0.5*	*thr = 0.8*	*thr = 0.3*	*thr = 0.5*	*thr = 0.8*
1	*30%*	*15%*	*6%*	*20%*	*15%*	*8%*
8	*6%*	*1.5%*	*2%*	*12.5%*	*6%*	*8%*
17	*40%*	*8%*	*6%*	*25%*	*7%*	*5%*

Another aspect that emerges from the analysis is that the Accuracy performance metric is not always the best solution to determine how important a feature set is. Regarding the values obtained for electrode 8, a big difference can be observed between the results of the 2 metrics, the F1 score having better statistics than accuracy regarding the performance in learning the characteristics of action potentials.

B. *Analysis of the Superlet with different order*

Following the results obtained from Table II, comparing the prediction values for the three orders and applying a threshold of 0.5, a significant difference can be identified. Thus, for the Superlet of order 5, with regard to the values of the two metrics, they are lower than the Superlet of order 1, respectively 2. This demonstrates the fact that, although the Superlet of order 5 offers a better resolution in the frequency domain, the temporal characteristics of the spikes play a crucial role in the process of spike sorting.

Table II. Comparative analysis of the difference in performance when training on the original and perturbed sets depending on the performance metric, correlation threshold (THR), and Superlet parameters (ORD, NCYC) on real data.

CHANNEL	THR	ORD	NCYC	ACCURACY	F1 SCORE
17	0.3	1	1.5	30%	25%
	0.5	1	1.5	17.5%	20%
	0.3	2	1.5	40%	25%
	0.5	2	1.5	8%	7%
	0.3	5	3	25%	15%
	0.5	5	3	3%	1.5%

C. *Ethical Statement*

The experiments conducted in this study strictly adhered to the ethical guidelines and regulations set forth by the European Communities Council Directive 86/609/EEC, as well as the directive 2010/63/EU of the European Parliament and Romanian Law 43/2014. These provide guidelines for the protection and ethical treatment of animals used in scientific research.

The experimental procedures and protocols were reviewed and approved by the Local Ethics Committee, with the approval number 3/CE/02.11.2018, and the National Sanitary and Veterinary Authority, with the approval number ANSVSA 147/04.12.2018. The experiments were conducted in accordance with the ethical guidelines outlined in the European directive, as well as the guidelines set by the Society for Neuroscience and the Romanian laws for the protection of animals.

To minimize the number of animals used and ensure their welfare, multiple datasets were collected over a period of 4 to 8 hours from each animal. This approach reflects a commitment to reducing the number of animals required for experimentation while still obtaining reliable and meaningful data.

V. CONCLUSION

In comparison with the Wavelet, respectively with the Superlet of order 5, it was proven that the Superlet transform of order 2, and the number of cycles 1.5, achieves a better performance regarding spike sorting. This aspect is due to the consideration that the 5th order transform loses its temporal precision, compared to the Wavelet and the 2nd order superlet. The latter seems to offer a sufficient precision both in time and frequency to enable a reliable identification of spikes.

Regarding the data permutation, we found that choosing a lower correlation threshold produces a larger difference in performance, while higher thresholds will result in a smaller difference in performance. This happens because a higher correlation threshold, on average, will lead to the formation of smaller feature sets and thus, their perturbation will have a smaller impact on performance.

The most important conclusion however pertains to the reason why the Superlet transform offers better separability of spike clusters, enabling a classifier to correctly trace the individual spikes to the neurons that generate them. In particular, both the frequency and temporal information help segregating different spike shapes. As figures 7, 8 and 12 indicate, information about spike characteristics is localized predominantly around the spike peak, extending upwards in frequency, and around a lower frequency component that carries a large fraction of the energy of the spike. Perturbations of this area have the most impact on the learning performance, indicating its relevance to the separability between the activity of different neurons. On the other hand, the amplitude peak of spikes and the temporal width of this peak also seems to be very informative. Indeed, spikes of excitatory neurons are usually wider and larger than spikes that arise in inhibitory neurons [9].

To conclude, we have explored why the Superlet transform provides useful features for spike sorting. We showed that it is able to isolate important time and frequency components that enable distinguishing spikes of different neurons from one another.

ACKNOWLEDGMENTS

The research leading to these results has received funding from: NO (Norway) Grants 2014-2021, under Project contract number 20/2020 (RO-NO-2019-0504), four grants from the Romanian National Authority for Scientific Research and Innovation, CNCS-UEFISCDI (codes PN-III-P1-1_1-TE-2021-0709, PN-III-P3-3.6-H2020-2020-0109, ERA-NET-FLAG-ERA-ModelDXConsciousness, and ERANET-NEURON-Unscrambly), and a H2020 grant funded by the European Commission (grant agreement 952096, NEUROTWIN).

REFERENCES

[1] H. G. Rey, C. Pedreira, and R. Quian Quiroga, "Past, present and future of spike sorting techniques," *Brain Res. Bull.*, vol. 119, pp. 106–117, Oct. 2015, doi: 10.1016/j.brainresbull.2015.04.007.

[2] D. B. Mureșan, R.-D. Ciure, E. R. Ardelean, V. V. Moca, R. C. Mureșan, and M. Dînșoreanu, "Spike sorting using Superlets: Evaluation of a novel feature space for the discrimination of neuronal spikes," in *2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sep. 2022, pp. 229–235. doi: 10.1109/ICCP56966.2022.10053955.

[3] V. V. Moca, H. Bârzan, A. Nagy-Dăbâcan, and R. C. Mureșan, "Time-frequency super-resolution with superlets," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41467-020-20539-9.

[4] H. Bârzan, A.-M. Ichim, V. V. Moca, and R. C. Mureșan, "Time-Frequency Representations of Brain Oscillations: Which One Is Better?," *Front. Neuroinformatics*, vol. 16, 2022

[5] M. O. İncetaş, "Image Interpolation Based on Spiking Neural Network Model," *Appl. Sci.*, vol. 13, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/app13042438.

[6] S. Fadnavis, "Image Interpolation Techniques in Digital Image Processing: An Overview," *Int. J. Eng. Res. Appl.*, vol. 4, pp. 2248–962270, Nov. 2014.

[7] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesth. Analg.*, vol. 126, no. 5, p. 1763, May 2018, doi: 10.1213/ANE.0000000000002864.

[8] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, Nov. 2018, doi: 10.1016/j.heliyon.2018.e00938.

[9] J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds, "Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4," Neuron, vol. 55, no. 1, pp. 131–141, Jul. 2007, doi: 10.1016/j.neuron.2007.06.018.