# Title

Sorting Overlapped Spikes using Superimposed Template Correlation Matching

By Srdjan Sumarac

# Significant Contributions

Designed and evaluated an algorithm to improve classification of non-isolated neural spikes.

# Poster

To add later

# Executive summary

A problem that is still present in modern day spike sorting is the presence of overlapping spikes. Resolving these waveforms at a higher accuracy would significantly improve the decoding of neural populations in the brain. In this paper, an add-on module approach that is capable of classifying and resolving overlapping spikes automatically at a higher accuracy is proposed. After constructing template waveforms from initial clustering of well-isolated spikes, a superimposed template generated from pair-wise combinations of those templates at all possible time shifts from each other is then implemented. Subsequently, overlapping waveforms were identified using a weighted absolute difference template matching method, and each spike waveform is assigned to their respective overlapping template label, and then automatically translated to the equivalent single-unit combination. This method was examined using simulated data from the SYNTH MONOTRODE dataset and the performance was evaluated using the Spike Forester accuracy metrics such as precision, recall and overall accuracy. It was found that the add-on module was able to resolve overlapping waveform at an average of 54% across the whole study, with various signal-to-noise ratios, an improvement of 4% from standard methods without the module. TALK ABOUT PRECISION AND RECALL TOO… As a result, this module is easily added as an extra step to improve standard low-dimensional clustering techniques such as k-means, dbscan, etc, in order to boost performance.

# Introduction

When recording extra-cellularly from multiple neurons, it is common to have action potential waveforms from one neuron altered by the action potentials of other neurons. This often results in waveform shapes and amplitudes that are significantly different from single-unit waveforms, thereby posing problems for spike sorting algorithms, especially when those algorithms are automated. This problem is exacerbated when using devices like electrode arrays (Normann et al., 1989), silicon electrodes (Anderson et al., 1989), and tetrodes (Wilson and McNaughton, 1993; Gray et al., 1995), which enable recording from large numbers of neurons simultaneously. Resolving these waveforms correctly into their constituent single units can be critically important for a range of analyses like response variability (Berry et al., 1997; de Ruyter van Steveninck et al., 1997), correlation (Maldonado et al., 2008; Cohen and Maunsell, 2009; Ecker et al., 2010), latencies (Keat et al., 2001; Gollisch and Meister, 2008), and information rates (Reich et al., 2000, 2001). Many attempts have been made to address this problem, but the resolution of overlapping waveforms is still not routinely performed in spike sorting. This could be due to limiting constraints of some of the available methods, some of which are described in the following paragraphs, or because of the high complexity of some other methods, which makes incorporating them into current spike sorting workflows difficult.

The most common approaches used to resolve overlapping waveforms in extracellular recordings create combinations of previously identified single-unit waveforms, usually referred to as templates, and then measure the similarity between the overlapping waveforms and the template combinations. These methods usually use the amplitude values of the waveforms to find the best match, although other features such as Fourier coefficients of the waveforms have also been used (Rinberg et al., 2003; Wang and Liang, 2005). Different techniques, such as different machine learning techniques, were used to compare the overlaps with the templates, such as support vector machines (SVM) (Ding and Yuan, 2008) and RELAX (Li and Stoica, 1996; Wang et al., 2009), a decoupled parameter estimation algorithm, that was used by Wang and Liang (2005). Also, Zhang et al. (2004) subtracted the templates from the overlaps, and used the similarity of the residue with a Gaussian distribution to find the best match, as they assumed the noise distribution was Gaussian. A similar method was used by Vargas-Irwin and Donoghue (2007). In all these techniques, a certain threshold needs to be determined to find the best matching template. Template-matching methods usually tend to be slow as well, because comparisons of all potential overlapping waveforms with all template combinations are needed. Recently, Adamos et al. (2010), addressing most of these issues, used a neural network to match overlapping waveforms with the templates generated by superimposing single-unit templates. However, it seems that the capability of this neural network in rejecting waveforms that do not belong to any of the identified single units, or combinations of those single units, remains to be investigated.

A very similar approach to Zhang et al. (2004) and Vargas-Irwin and Donoghue (2007) was taken by Prentice et al. (2011) and Pillow et al. (2013), in which they also subtracted the best matching spikes from the recorded signal until the residue was indistinguishable from noise. The difference was that these methods used greedy algorithms to find the match instead of using a brute-force search. This meant they will scale better. However, both these methods needed to make assumptions about factors (such as the distribution of noise or spike trains) that are simplified. For example, Prentice et al. (2011) assumed a Poison distribution for the firing of the cells, which is reasonable, but may not be applicable in a number of cases (Berry et al., 1997; Berry and Meister, 1998). As mentioned by both these studies, adding more constraints to make these assumptions more similar to real recordings may make these methods computationally expensive.

Other methods that do not involve template matching concentrated on extracting more robust features other than amplitude to cluster overlapping waveforms with single-unit waveforms. Among these methods, Hulata et al. (2002) and Quiroga et al. (2004) represented the waveforms using wavelet coefficients, and used k-means clustering and superparamagnetic clustering, respectively. However, they appear to have trouble resolving overlapping waveforms resulting from the near simultaneous firing of multiple neurons.

Another group of methods aimed to decompose the overlaps into their components. For example, in Oweiss and Anderson (2007), the overlapping waveforms were decomposed using the discrete wavelet packet transform, Takahashi et al. (2003) applied the independent components analysis (ICA) technique, and Franke et al. (2010, 2015) used a set of linear filters to decompose the overlapping waveforms. For these methods, like other methods discussed earlier, one needs to set a threshold to measure the similarity. Franke et al. (2010, 2015) used an analytical method to find the threshold, but it seems that to use this method, it is necessary to assume a Gaussian distribution for noise, as well as the distribution of single-unit spikes. Moreover, the other drawback of the method described in Takahashi et al. (2003) and Franke et al. (2010, 2015) is that two single units with different amplitudes but similar amplitude patterns cannot be distinguished from each other, although this could be advantageous in dealing with bursting cells. Ekanadham et al. (2014), who used Continuous Basis Pursuit in order to estimate the most probable spike patterns given the observed recording, also suffered from the same drawback, although their method has the advantage that it can be scaled for sorting multi-electrode array recordings.

More recently, methods developed for use with high density multi-electrode arrays, with much smaller spacing between electrodes, have taken advantage of the fact that the activity of one neuron appears on several different electrodes, and the spatial information provided by these arrays can be used to resolve overlapping waveforms (Marre et al., 2012; Pachitariu et al., 2016; Rossant et al., 2016; Yger et al., 2017; see review by Lefebvre et al., 2017). However, a lot of recordings in larger animals, and in human subjects, are still performed with electrode arrays with electrode spacing larger than 200 μm (e.g., Blackrock Utah Array, Microprobe Floating Microelectrode Arrays, etc.). These electrode separations greatly reduce the likelihood of multiple electrodes recording the activity of one neuron, and thus negates some of the advantages that these methods provide.

In order to eliminate some of these difficulties, we propose a partially automated method to resolve overlapping waveforms based on template matching. Our method uses KlustaKwik to perform the clustering, and the MClust software package to visualize and inspect the results of the spike sorting. As KlustaKwik is among the most popular automated clustering methods used for spike sorting (Wild et al., 2012), and MClust is a widely used package for spike sorting, applying our method to currently sorted datasets may be less problematic compared to some other methods. Our method has been designed to be added as an extra step to spike sorting routines currently in use for single electrode and tetrode recordings, and has the advantage that it can be parallelized in order to decrease the computational load. In our method, all steps, except the single-unit template selection, are automated and unsupervised. We show that despite the simplicity of our automated approach, it can successfully produce spike trains with fairly high accuracy in simulated data.

# Literature Review

# Overview

# Detailed Discussions

# Evaluation

# Results

# Future Work

Try to implement overlapping techniques to multi-channel recordings due to the improvement of dense electrode arrays.

# Conclusion

# References

# Appendices