UNIVERSITY OF CALIFORNIA,
IRVINE

Action Potentials and Waves: A Short Story on Electrophysiological Signal Processing

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Agnieszka A. F. Szymanska

Dissertation Committee:
Associate Professor Zoran Nenadic, Chair
Professor Frithjof Kruggel
Assistant Professor Beth Lopour

2017

DEDICATION

To Professor Richard Haskell who taught me how to be a scientist.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation owes an inexpressible debt of gratitude to the exacting standards, enduring support, and guidance of my advisor, mentor, and chair of my committee, Dr. Zoran Nenadic. His dedication to rigorous and thorough science elevated the work in this dissertation, and he has my sincere and heartfelt thanks.

Dr. Frithjof Kruggel and Dr. Beth Lopour also served on the committee, providing warm guidance along the way as well as challenging and insightful questions to guide me on the right track. They both have my thanks for their diligent efforts.

It has been a privilege to share time and space with Dr. Enrico Gratton - he has my thanks for his wise and thoughtful counsel in our collaborations and beyond. Thanks also to Dr. Yuji Ikegaya, who stewarded my work in Japan. He not only guided me through the many intricacies of Japanese academic life throughout our collaborations, but greatly deepened my expertise in neuropharmacology. My thanks are also extended to Dr. Yama Akbari, one of the finest academics I have had the pleasure to work with, who did me the honor of collaborating on exciting clinically relevant projects that are sure to lead to numerous publications.

A dissertation like this owes much to labmates. Chang Won Lee, Christine King, Po Wang, Colin McCrimmon, Michael Doty, Marta Alcala Alvarez, and Kristoffer Boyle Miller have been a constant presence, providing humor, warmth, encouragement, and support.

I am thankful for my colleagues, the fabric of university life, for giving me such a vibrant and rich experience: Alba Alfonso Garcia, Rupsa Datta, Tim Smith, Robert Rague, Scot Parker, Chris Heylman, Afsheen Bazrafkan, Rachel Smith, Siavash Ahrar, Yasemin Serigul-Klijin, and Dail Chapman.

Behind every great endeavor is a network of support, and it is a humbling privilege to be surrounded by people I admire. Thank you to my friends, especially Paul Roach, who have all taken an interest in my work and done an admirable job learning about graduate academia. Thank you to my siblings, Asia and Olek for their patience during my extended work on my dissertation. Thank you to my parents, Anna and Artur, for their constant support, and encouragement. And, at the center of it all, my greatest thanks to my devoted and steadfast partner, Daniel.

Thank you to IEEE and Elsevier publishers for their permission to incorporate the work I've published in their journals into my dissertation.

This work was funded by the National Science Foundation (1056105, and OISE-13 7307954), the National Institute of Health (P41-GM103540, P50-GM076516, and 1UH2TR000481-01), and the Japanese Society for the Promotion of the Sciences (SP15054).

CURRICULUM VITAE

Agnieszka A. F. Szymanska

EDUCATION

Doctor of Philosophy in Biomedical Engineering **2017**
University of California, Irvine *Irvine, CA*

Master of Science in Biomedical Engineering **2014**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Physics **2011**
Harvey Mudd College *Claremont, CA*

RESEARCH EXPERIENCE

Graduate Research Assistant **2011–2017**
University of California, Irvine *Irvine, CA*

Student Researcher **2008**
Harvey Mudd College *Claremont, CA*

TEACHING EXPERIENCE

Teaching Assistant **2016**
University of California, Irvine *Irvine, CA*

Teaching Assistant **2008–2011**
Harvey Mudd College *Claremont, CA*

REFEREED JOURNAL PUBLICATIONS

A.F. Szymanska, M. Alcala Alvarez, M. Farahabadi, A. Bazrafkan, Z. Nenadic, and Y. Akbari. Effects of increasing anesthesia on spontaneous neuronal activity during burst suppression in M1 of rat cortex **2017**
PLOS ONE (*In Progress*)

A.F. Szymanska, C. Heylman, R. Datta, E. Gratton, and Z. Nenadic. Automated detection and analysis of depolarization events in human cardiomyocytes using MaDEC **2016**
Computers in Biology and Medicine

A.F. Szymanska, C. Kobayashi, H. Norimoto, T. Ishikawa, Y. Ikegaya, and Z. Nenadic. Accurate Detection of Low Signal-to-Noise Ratio Neuronal Calcium Transient Waves Using a Matched Filter **2016**
Journal of Neuroscience Methods

C. Kobayashi, K. Okamoto, Y. Mochizuki, Hi. Urakubo, T. Ishikawa, K. Funayama, N. Imamura-Hitora, A. F. Szymanska, H. Hioki, S. Ishii, Y. Ikegaya. Dendrites Screen Presynaptic Neuron Ensembles **2017**
Journal of Physiology (*Under Review*)

REFEREED CONFERENCE PUBLICATIONS

A.F. Szymanska, M. Doty, K.V. Scannell, and Z. Nenadic. A supervised multi-sensor matched filter for the detection of extracellular action potentials **2014**
Proc. of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

K. Scannell, A.F. Szymanska, and Z. Nenadic. Intrinsic dimensionality of extracellular action potentials **2014**
Proc. of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

A. Szymanska, A. Hajirasooliha, and Z. Nenadic. Source location as feature for the classification of multi-sensor extracellular action potentials **2013**
Proc. of the 6th Annual International IEEE EMBS Conference on Neural Engineering
(Symposium on Problems at the Neural Interface - Best Student Paper Winner)

A. Szymanska and Z. Nenadic. Wavelet-approximated generalized matched filter for the detection of multisensor extracellular action potentials **2013**
Proc. of the 6th Annual International IEEE EMBS Conference on Neural Engineering

C.W. Lee, A. Szymanska, Y. Ikegaya, and Z. Nenadic. The accuracy and precision of signal source localization with tetrodes **2013**
Proc. of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

A.H. Do, P.T. Wang, C.E. King, A. Schombs, J.J. Lin, M. Sazgar, F.P.K. Hsu, S.J. Shaw, D.E. Millett, C.Y. Liu, A. Szymanska, L.A. Chui, and Z. Nenadic. Sensitivity and specificity of upper extremity movements decoded from electrocorticogram **2013**
Proc. of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

CONFERENCE ABSTRACTS

A.F. Szymanska, M. Alcala Alvarez, M. Farahabadi, A. Bazrafkan, Z. Nenadic, and Y. Akbari. Different isoflurane levels have distinct effects on cross-hemisphere functional connectivity during unconsciousness **2016**
Society for Neuroscience, San Diego, CA, USA

A. F. Szymanska. Matched filter for multi-unit calcium event detection **2015**
Multidisciplinary Science Forum of the US JSPS Alumni Association, Sacramento, CA, USA

INVITED TALKS

A.F. Szymanska. Matched filter for multi-unit calcium event detection **2015**
Multidisciplinary Science Forum of the US JSPS Alumni Association, Sacramento, CA, USA

A.F. Szymanska. Accurate detection of low SNR calcium transient waves using a matched filter **2015**
Biomedical Engineering Seminar Series, University of California Irvine, Irvine, CA, USA

A.F. Szymanska, M. Doty, K. Scannell, Z. Nenadic. A supervised multi-sensor matched filter for the detection of extracellular action potentials **2014**
36th Annual International Conf. of the IEEE EMBS, San Diego, CA, USA

A. Szymanska, A. Hajirasooliha, Z. Nenadic. Source location as a feature for the classification of multi-sensor extracellular action potentials **2013**
Pre-conference symposium on the neural interface, 6th International IEEE EMBS Conference on Neural Engineering, San Diego CA

A.H. Do, P.T. Wang, C.E. King, A. Schombs, J.J. Lin, M. Sazgar, F.P.K Hsu, S.J. Shaw, D.E. Millett, C.Y. Liu, A. Szymanska, L.A. Chui, Z. Nenadic. Sensitivity and Specificity of Upper Extremity Movements Decoded from Electrocorticogram **2013**
35th Annual International Conf. of the IEEE EMBS, Osaka, Japan

SOFTWARE

MMiCE `http://sites.uci.edu/aggies/downloads/`
The MMiCE detector is a software package designed to detect calcium events from multiple calcium imaged neuronal somas, neuronal dendritic spines, or astrocytes. The input should be a time series of spectral intensity, for each pre-specified region of interest (such as somas or dendritic spines), as a function of time. The software is used through a GUI. Below, you can download sample data, sample training as well as tutorials describing how to use and train the detector.

MaDEC `http://sites.uci.edu/aggies/downloads/`

MaDEC is a software package designed to detect, quantify, and compare depolarization events from voltage sensitive dye imaged beating cardiomyocytes. The input should be a time series of spectral intensity, for each pre-specified region of interest (such as a line-scan etc.), as a function of time. The software is used through a GUI.

ABSTRACT OF THE DISSERTATION

Action Potentials and Waves: A Short Story on Electrophysiological Signal Processing

By

Agnieszka A. F. Szymanska

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2017

Associate Professor Zoran Nenadic, Chair

Electrophysiological signal processing is a broad, complex and growing field. Universally, the first steps of any signal analysis are detection and classification. Here we present a flexible matched filter designed to detect spikes from various biological data types, as well as two statistically based approaches for spike classification. We then apply these developed tools to study the effects of deep anesthesia on neuronal network dynamics.

The matched filter was implemented for three different applications: detecting action potentials (APs) from multi-sensor extracellular recordings, detecting depolarization events (DEs) from voltage sensitive dye (VSD) imaged cardiomyocytes, and detecting calcium events (CEs) from calcium imaged neuronal somas as well as dendritic spines. In the case of AP detection, the filter performed with average TP and FP rates of 85% and 17%, respectively, as compared with manual detection. In the case of DE detection, the filter performed with TP and FP rates of 98%-100% and 1%-2%, respectively, as compared with manual detection. In the case of CE detection, the filter performed with TP and FP rates of 100% and 2%, as verified by patched clamp recordings. It also reached perfect performance on simulated data at SNR ≥ 2, and TP and FP rates of 98% and 7% on simulated data with SNR $= 0.2$. Overall, the presented matched filters can accurately detect spikes from various kinds of biological data.

The classification problems explored here include AP (or spike) sorting, as well as DE classification across different drug administrations. In the case of spike sorting, the MUSIC algorithm was used to extract neuronal source locations from multi-sensor extracellular AP recordings. The source locations were then used as classification features. This approach was able to reliably classify tetrode (4 channel) and heptode (7 channel) recorded APs. For DE classification, salient DE features were extracted and then compared across drug treatments using a Kolmogorov-Smirnov test. The drug treated cells were consistently statistically distinguishable from controls, even at SNR ≤ 3. The method could also distinguish cells immediately after drug administration from cells 10 min after drug administration. Overall, both methods success makes them valuable tools for studying neuronal networks as well as cardiomyocyte drug assays, respectively.

Finally, the matched filter for AP detection as well as the MUSIC-based AP classification scheme were applied to *in vivo* heptode data collected from M1 of the right hemisphere of anesthetized rats. The rats were also fitted with two electrocorticography (ECoG) electrodes over M1 and V1 of the left hemisphere. Data was collected at three monotonically increasing isoflurane anesthesia levels to assess the effects of increasing anesthesia on neuronal network dynamics during a burst suppression state. To our knowledge, this is the first such study. We found that higher anesthesia led to higher AP frequency, no change in the number of active single units, and increased cross-hemisphere functional connectivity. Additionally, all APs were restricted to ECoG bursts, with no APs occurring during suppressed ECoG states.

Chapter 1

Introduction

The work presented in this dissertation deals with a very broad topic: electrophysiological signal processing. Strictly speaking some of the signals, such as neuronal calcium transient waves, are actually only a proxy for electrophysiological activity. Nonetheless the ultimate goal is to develop processing tools to study the electrical activity of live tissues.

The most obvious electrically active cells are, of course, neurons. Although many techniques have been employed in the study of higher-order neural function, such as single unit patch clamping, extracellular recording, electroencephalography (EEG), pharmacological manipulation, calcium imaging, optogenetics, and functional magnetic resonance imaging (fMRI), they can all be stripped down to the study of a single phenomenon – neuronal spike trains [10]. Transient voltage pulses traveling down axons, commonly known as action potentials (APs), are the main form of neural communication, and computation [45, 17, 49, 19]. The study of APs and how they propagate to convey information through neuronal populations is therefore fundamental to neuroscience. In this work we explore several different kinds of neurological recordings: multi-sensor extracellular recording, which records electrical signals from several neurons in the vicinity of the electrode, functional multi-neuron calcium imaging (fMCI),

which records the calcium transient wave activity of several hundred adjacent neurons, and electrocorticogram (ECoG) which records the global electrical signals of a population of cortical neurons located under a roughly 1mm surface area electrode.

Another type of electrically active cells explored in this dissertation are cardiomyocytes. Although it does not resemble a neuronal AP, the electrical activity of cardiomyocytes is also often referred to as an action potential. However, to avoid confusion we will refer to cardiomyocyte electrical activity as depolarization events (DEs). There are also many ways to record cardiomyocyte DEs. Patch clamping offers precise measurement of transmembrane voltages of single cells [5, 37, 111]; microelectrode arrays offer a minimally invasive platform for assessing many extracellular potentials [51, 96]; calcium reporters such as Fluo-4 [75, 84], genetically encoded sensors such as GCaMP [38, 44, 95], and voltage sensitive dyes (VSDs) [35, 67] offer minimally invasive, fluorescence-based detection of DEs. VSD based fluorescence intensity measurements are the cardiomyocyte recordings presented in this work.

Once the signals are collected, signal processing can begin. Universally, the first step in any analysis is detecting the signal of interest, followed most often by picking out salient features of the signal and then often comparing different features or experimental paradigms. These basic steps, detection and classification, are often called pre-processing, because they must be performed before more complex comparisons of studies or conditions can be performed. Although often mentioned in passing, these steps are often the most computationally intensive and complex signal processing performed on the data. In this work we will discuss a machine learning approach to signal detection, applied to three different data types - extracellular recording from neurons, calcium imaging of neurons, and voltage sensitive dye imaging of caridomyocytes - as well as a two statistical approaches to classification, applied to two different data types - extracellular recording from neurons, and voltage sensitive dye imaging of caridomyocytes. The dissertation ends with applying all of these techniques in the process of discovery to investigate the effects of deep anesthesia both on local neural

network dynamics and on the local network relationship with cross-hemisphere global neural activity.

For the benefit of the reader the following sections present executive summaries of the chapters.

1.1 Executive Summary: Detection

As mentioned before, detection is a crucial component of any biophysical signal processing algorithm. Electrophysiological data is often riddled with biological noise which is both correlated with and statistically similar to the signal of interest, making traditional detection tools from other fields inapplicable. With that in mind we have adopted a machine-learning approach and developed a flexible matched filter to detect spikes from various types of biological data.

1.1.1 Matched Filter Design

Although the matched filters implemented throughout this work have analyzed very different kinds of data, they all share a common base design. Assuming Gaussian noise statistics, we can express a generalized matched filter (GMF) [46] as

$$S(x) = s\Sigma^{-1}x^T \quad , \quad \begin{array}{lll} H_1 & \text{if} & S(x) > \gamma \\ H_0 & \text{if} & S(x) < \gamma \end{array} \tag{1.1}$$

where $x \in \mathbb{R}^{1\times CN}$ is the row vector form of the C-sensor, N-sample input matrix, $s \in \mathbb{R}^{1\times CN}$ is the row vector form of the "spike" (i.e. the template), $\Sigma \in \mathbb{R}^{CN\times CN}$ is the spatio-temporal noise covariance matrix of the row vector form of the noise, γ is the threshold, H_0 is the null

hypothesis, and H_1 is the alternative hypothesis. The filter is convolved with a C-sensor, T-sample ($T \gg N$) time series to detect spikes at a given threshold γ.

1.1.2 Applications

The matched filter presented in this work has been applied to three different data types: extracellular recordings of neurons, voltage sensitive dye imaging of cardiomyocytes, and calcium imaging of neurons - all electrically active cells that exhibit a "spike" as the signal of interest.

1.1.2.1 Multi-sensor Extracellular Action Potential Detection

The work presented in this section is adapted from Szymanska et al. 2013 [99]. Briefly, a matched filter was designed and implemented to detect APs from multi-sensor extracellular recordings. The detector was tested on tetrode data from a locust antennal lobe and assessed against three trained analysts. 25 APs and noise samples were selected manually from the data and used for training. To reduce complexity, the filter assumed that the underlying noise in the data was spatially white. The detector performed with average TP and FP rates of 84.62% and 16.63% respectively. This high level of performance indicates the algorithm is suitable for widespread use.

Methods 20 seconds of data from four sensors of a planar microelectrode placed in an adult locust's antennal lobe were used for this study [78, 79]. Half of the collected data was used for training (training data), and the remaining half was used for further analysis (test data). Three trained analysts independently tagged all of the spikes in the test data. Twenty five APs (2 ms each; $N = 30$, $C = 4$), and 25 noise samples ($\sim$ 30 ms each) were manually selected from the training data and used to generate the matched template, s, and the noise

covariance matrix, Σ, respectively. The noise statistics were assumed to be spatially white. Detected APs, given 40 incrementally increasing thresholds, were compared against those tagged by each trained analyst, with the analyst acting as the "ground-truth". The TP and FP rates at each threshold were then used to generate receiver operating characteristic (ROC) curves for each analyst.

Results The detector performed well compared with all three analysts. The optimal threshold for each analyst was determined by minimizing the distance between their respective ROC curve and theoretically perfect performance (100% TP, 0% FP). At the optimal threshold, the detector performed with a TP rate of 90.79% and an FP rate of 20.66% compared with Analyst 1, a TP rate of 80.00% and an FP rate of 10.31% compared with Analyst 2, and a TP rate of 83.06% and an FP rate of 18.91% compared with Analyst 3. The detector's average TP and FP rates were 84.62% and 16.63%, respectively.

Conclusion The multi-sensor matched filter developed in this study was assessed against three trained analysts and performed with average TP and FP rates of 84.62% and 16.63%, respectively. The detector's performance presents it as a great candidate for wide-spread use as one of the only multi-sensor detectors of its kind. Furthermore, the algorithm's performance is likely under estimated as a vast majority of the spikes identified as FPs for Analysts 1 and 2 were actually identified as TPs by Analyst 3. Because the analysts were proven to be inconsistent and unreliable, we believe that the detector would perform even better if compared against the ground-truth instead of trained analysts.

1.1.2.2 MaD

The work presented here is adapted from Szymanska et al. 2016 [103]. Briefly, a matched filter for depolarization event detection (MaD) was designed and implemented to detect DEs

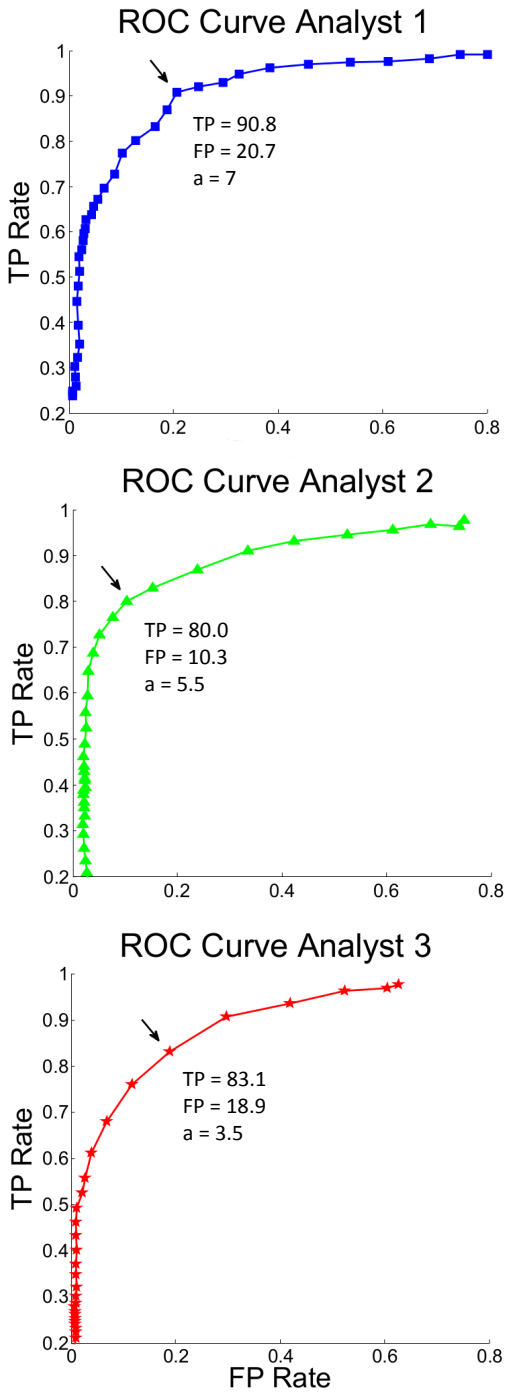

Figure 1.1: ROC Curves for all three analysts. TP and FP pairs were collected for 40 incrementally increasing thresholds, a, for each analyst. The black arrows point to the detector's optimal performance for each analyst, with the appropriate parameters listed below.

from voltage-sensitive dye (VSD) imaged cardiomyocytes. We used 2-photon microscopy of fluorescent VSDs to capture the membrane voltage of actively beating human induced

pluripotent stem cell-derived cardiomyocytes (hiPS-CMs). The efficacy of the MaD detector was quantified by comparing detection results against manual DE detection by expert analysts. MaD accurately detected DEs with true positive rates of 98-100% and false positive rates of 1-2%, at signal-to-noise ratios (SNRs) of 5 and above, given three hiPS-CM drug treatments: propranolol, isoproterenol, and control.

Methods Prepared hiPS-CMs began spontaneously beating on approximately Days 12-15 of culture. On Day 33, the cells were stained with VSD, treated with β-adrenergic drugs, and then imaged. One culture was treated with propranolol, two cultures were treated with isoproterenol, and two cultures were left untreated as controls. Two-photon microscopy line scan data was then collected for all of the cultures. The drug treated cultures were imaged immediately after addition of drugs (less than 60 sec of exposure) and again 10 min or 15 min after addition to ensure complete exposure. This accounts for a total of 7 data traces, from 5 cell cultures. The imaging data was then processed to extract fluorescence intensity traces along a given cell membrane, and filtered to remove photobleaching artifacts. Intensity traces, X, for all drug and control conditions were then plotted in Matlab, and three trained human analysts independently identified DE peak times from each trace. Unanimous DEs between all 3 analysts were used as the "ground-truth" for assessing MaD performance.

The detector was trained under three detection conditions: control, isoproterenol exposure, and propranolol exposure. Twenty DEs ($N = 800 \simeq 0.48$ s), and 20 noise samples (~ 30 ms each) were manually selected from the data and used to generate the matched template, s, and the noise covariance matrix, Σ, respectively. The detector was tested using the full noise covariance to estimate Σ.

Detected DEs, at 45 incrementally increasing thresholds, were compared against the analyst estimated "ground-truth", and TP and FP rates at each threshold were then used to generate ROC curves for each detection case.

Results The threshold at which detector performance is closest to TP = 100% and FP = 0% is the optimal threshold and constitutes the best detector performance for that detection case. All following performance metrics are presented as TP or FP Rate [95% confidence interval].

The best performance for the control case was TP = 97.92 [92.15, 100.00]% and FP = 1.75 [0.00, 5.26]%. The best performance for the isoproterenol case was TP = 100.00 [100.00, 100.00]% and FP = 0.78 [0.00, 2.33]%, and the best performance for the propranolol case was TP = 98.81 [96.43, 100.00]% and FP = 1.19 [0.00, 3.57]%. Overall, the MaD detector performed on par with human analysts and accurately identified DEs at SNR levels of 5 and above.

In order to better test the MaD detector in low-SNR environments we also tested MaD on a low-SNR control data set (SNR = 3.19), as well as a low-SNR isoproterenol exposure data set (SNR = 0.65). Performance decreased in the low-SNR detection cases. The best performance for the low-SNR control case was TP = 72.37 [61.84, 81.58]% and FP = 16.67 [7.58, 25.76]%, and the best performance for the low-SNR isoproterenol case was TP = 65.79 [49.99, 81.59]% and FP = 13.79 [0.44, 27.14]%.

Conclusion The MaD detector presented here provides a tool that is specifically designed for automatic detection of VSD imaged cardiomyocyte DEs. It has proven to work exceptionally well, especially on data with SNR $\geq$ 5, with TP rates ranging from 98-100% and FP rates ranging from 1-2%. Overall the MaD detector is a useful new tool for the study of cardiomyocyte electrophysiology. Combined with the use of voltage-sensitive dyes, it allows for non-invasive, image-based, and automated detection of cardiac DEs.

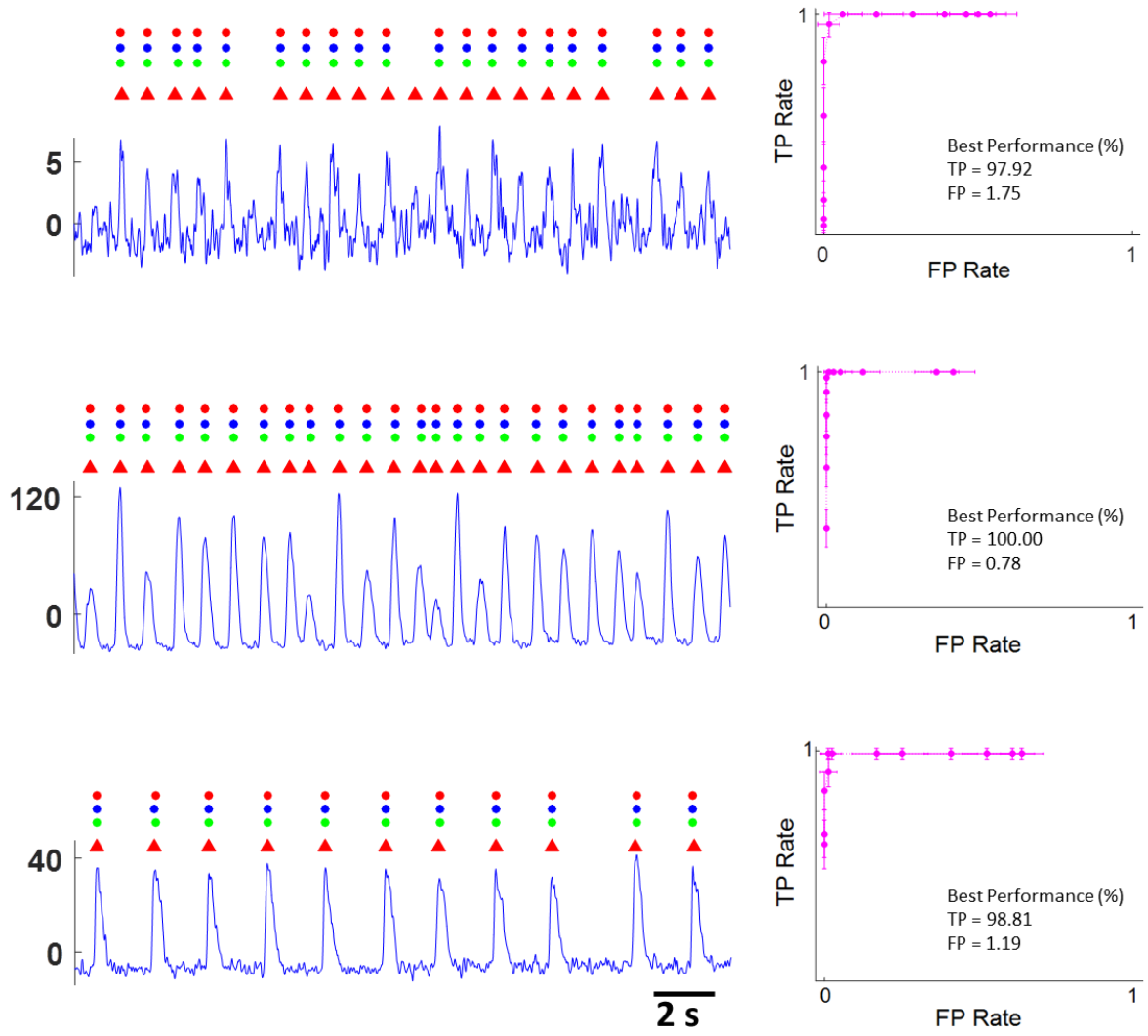

Figure 1.2: MaD Performance. The figure consists of three pre-processed intensity traces from the control (*Top*), isoproterenol (*Middle*), and propranolol (*Bottom*) drug treatments. The circles above each data trace represent the DEs identified by analysts 1-3 (bottom to top respectively). The triangles above each data trace represent MaD detected DEs, using the optimal thresholds identified in the corresponding ROC curves to the right. The optimal threshold is identified as the one resulting in the detector performance closest to TP = 100% and FP = 0%. The error bars represent 95% confidence intervals. The optimal performance for each drug treatment case is presented under each ROC curve. The MaD detector performed with a TP rate of 98-100% and a FP rate of 1-2% for all 3 drug treatments.

1.1.2.3 MMiCE

The work presented here is adapted from Szymanska et al. 2016 [104]. Briefly, we developed a Matched filter for Multi-unit Calcium Event (MMiCE) detection to extract calcium events (CEs) from fluorescence intensity traces of simulated and experimentally recorded neuronal calcium imaging data. MMiCE performed with a TP rate of 98.27% (with a 95% confidence interval of [96.89, 99.65]%), and an FP rate of 6.59% (with a 95% confidence interval of [4.03, 9.15]%) on simulated data with SNR 0.2. It reached perfect performance on simulated data with SNR $\geq$ 2. MMiCE performance on real data was also very promising (TP = 100 [100.00, 100.00]%, FP = 2.04 [0.00, 6.14]%). Overall, the MMiCE detector performed exceptionally well on both simulated data at SNR as low as 0.2, as well as experimentally recorded neuronal calcium imaging data. The MMiCE detector is accurate, reliable, and well suited for wide-spread use.

Methods MMiCE was tested under 3 paradigms: (1) simulated somatic calcium imaging data (ground-truth available), (2) experimentally recorded simultaneous somatic fMCI and patch-clamp data (ground-truth available), as well as (3) experimentally recorded somatic fMCI and dendritic spine fMCI data (no ground-truth available). In this summary we will concentrate on paradigms (1) and (2) as they have ground-truth available for quantitative analysis.

For the simultaneous somatic fMCI and patch-clam (fMCI-PC) recordings *ex vivo* organotypic slice cultures were prepared from the hippocampus of 7 day old Wistar/ST rats [105]. These cultures were incubated in an OGB1 dye solution, washed, and mounted in a recording chamber. CA3 pyramidal neurons selected for recording were then voltage-clamped at 0 mV. One data trace measuring spontaneous activity from a CA3 pyramidal neuron was acquired for the fMCI-PC recordings.

The simulated fMCI data was generated from 3 somatic fMCI data sets measuring spontaneous activity from CA1 pyramidal neurons from acute mouse brain slices. 400 μm horizontal hippocampal slice of 3 week old C57Bl/6J mice [108, 74] were locally loaded with an OGB1 dye solution into the CA1 stratum pyramidale using a micro pipette.

Fluorophores for all of the collected data were excited at 488 nm with a laser diode and visualized using a 507 nm long-pass emission filter. Regions of interest (ROIs) for all fMCI data sets were identified manually using a custom software [40]. The average fluorescence for each ROI was then calculated and used to determine the change in fluorescence $\Delta F/F$ using a 2 second normalization epoch.

To generate the simulated data, 4 distinct CE shapes were identified manually from the somatic fMCI data sets. 20 examples of each shape were averaged, low-pass filtered at 5 Hz, and normalized to construct 4 distinct CE templates. Similarly, 80 samples of varying length noise-only data were manually selected from the somatic fMCI data sets. Each sample was normalized to be zero mean and unit variance. The noise samples were then all concatenated into a single time series and the four CE templates were superimposed with the noise at random time points such that no spikes overlapped, and there were 100 CEs of each spike shape for a total of 400 CEs. The fluorescence traces were simulated at SNR levels of 0.2, 0.5, 1, 2, 4, 6, 9, and 14, where SNR 1 indicates that the signal has the same power as the noise.

To train the MMiCE detector for the simulated data, 20 high-SNR CEs ($N = 1.6$-s) and 20 noise-only samples of various lengths, from multiple different ROIs, were identified manually and used to generate s and Σ, respectively. For the fMCI-PC data, 20 noise-only samples of various lengths were identified manually to generate Σ, however due to a limited number CEs available in the single fMCI-PC data trace acquired, only 10 CEs ($N = 1.6$-s) were selected for generating s. All data selected for training were omitted from the performance analysis.

Detected CEs, at 100 incrementally increasing thresholds, were compared against the ground-truth, and TP and FP rates at each threshold were then used to generate ROC curves for each paradigm.

Results The MMiCE detector performed exceptionally well on the simulated data, even at SNR levels as low as 0.2. Performance metrics are listed as TP or FP Rate [95% confidence interval]. The MMiCE detector was first applied in its simplest form, assuming WGN statistics when estimating Σ. Under this assumption, the MMiCE detector achieved perfect performance (TP = 100 [100.00, 100.00]%, FP = 0 [0.00, 0.00]%) for SNR $\geq$ 2. At SNR 1 the performance was slightly affected (TP = 100 [100.00, 100.00]%, FP = 0.29 [0.00, 0.86]%), with a further dip in performance at SNR 0.5 (TP = 97.98 [96.84, 99.12]%, FP = 0.88 [0.00, 1.87]%) and SNR 0.2 (TP = 90.46 [87.35, 93.57])%, FP = 12.81 [9.34, 16.28]%). To see if even better outcomes could be achieved for SNR 0.2 - 2, the MMiCE detector was then applied using the full noise covariance. For simulated data sets of SNR 1 and SNR 2, performance was not affected. Performance was improved for SNR 0.5 (TP = 99.71 [99.14, 100.00]%, and FP = 1.15 [0.03, 2.27]%), and for SNR 0.2 (TP = 98.27 [96.89, 99.65]%, and FP = 6.59 [4.03, 9.15]%).

As in the simulated data case, the fMCI-PC data trace was filtered using the MMiCE detector under the WGN assumption first and achieved a performance of TP = 95.83 [89.97, 100.00]% and FP = 2.13 [0.00, 6.41]%, which constitutes 3 errors. The performance improved using the full noise covariance with TP = 100.00 [100.00, 100.00]% and FP = 2.04 [0.00, 6.14]%, constituting 1 error. Overall, the MMiCE detector did exceptionally well in CE detection from somatic fMCI data, as verified via patch-clamp recording.

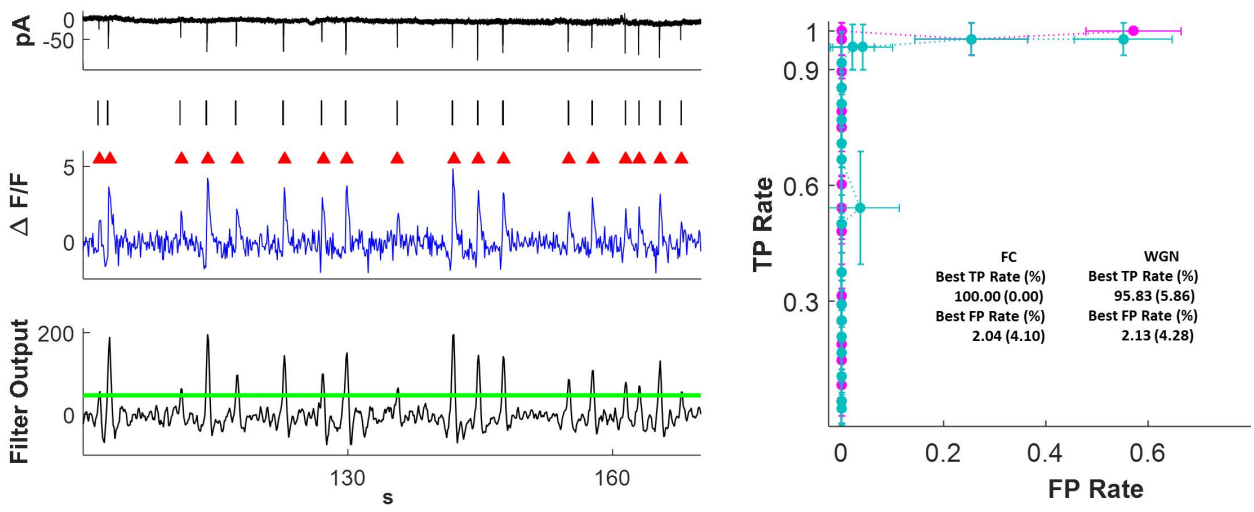

Figure 1.3: MMiCE detector performance and fMCI-PC data. (*Left*) The top shows the ground-truth patch-clamp recordings. In the middle is the corresponding $\Delta F/F$ data with MMiCE CE peaks detected at the optimal threshold marked by red triangles. The spike-train derived from the patch-clamp recordings is shown above the detected CEs for easy comparison. The bottom trace shows the filter output using the full noise covariance and the optimal detection threshold as determined by the ROC curves presented on the *Right*. (*Right*) ROC curves for the fMCI-PC data shown on the *Left*, across 100 incrementally increasing thresholds. The two curves depict the MMiCE detector's performance under the WGN assumption (cyan) and when using the full noise covariance (FC, magenta). The error bars represent 95% confidence intervals. Optimal performances with 95% confidence errors in parenthesis are presented under the curves.

Conclusion The MMiCE detector was designed to identify CEs in low-SNR environments, and tested on both simulated and experimentally recorded fMCI data. This time ground-truth data was available and the detector performed exceptionally well in both cases.

The simulated data was constructed from noise segments and CEs captured from real neuronal somatic fMCI recordings and varied from SNR 0.2 to SNR 14. MMiCE reached perfect performance at SNR 2 and above. Even at SNR 0.2 the MMiCE detector reached a TP Rate of 98.27 [96.89, 99.65]% and a FP Rate of only 6.59 [4.03, 9.15]%. The MMiCE detector's performance was also tested on simultaneously recorded somatic fMCI and patch-clamp data and achieved a TP Rate of 100.00 [100.00, 100.00]% and a FP Rate of = 2.04 [0.00, 6.14]%, which constitutes only 1 error. This high performance level was on par with and in some

cases exceeded that shown by existing methods [47, 88, 66, 119, 27], and did so while being tested at SNR levels as low as 0.2, well below those used in previous studies (SNR $\simeq$ 5 - 10). Overall, these results indicate the MMiCE detector is a valuable tool for detecting low-SNR neuronal CEs from imaging data.

1.1.3 Conclusion

In this section we presented three matched filters, all based on the same initial design intended to find spikes within time series data. The first filter was adapted to detect APs from multi-sensor extracellular recordings. It performed well on *in vivo* tetrode data and was able to achieve an average performance of TP Rate = 85% and FP Rate = 17% as compared with manual detection by 3 trained analysts. The second filter, the MaD detector, was adapted to detect DEs from VSD imaged cardiomyocytes. It performed well on fluorescent line scan data from *in vitro* hiPC-CMs with SNR $\geq$ 5. The MaD detector achieve TP rates of 98-100% and FP rates of 1-2% as compared with unanimous DEs manually detected by 4 trained analysts. The final filter, MMiCE, was adapted to detected CEs from fMCI data of neuronal somas and dendritic spines. The detector performed exceptionally well on simulated somatic fMCI data, with perfect performance at SNR $\geq$ 2, and TP Rate = 98% and FP Rate = 7% on simulated data with SNR = 0.2. Furthermore, the MMiCE detector reached a performance of TP Rate = 100% and FP Rate = 2% on experimentally recorded *in vitro* somatic fMCI data, as verified by patched clamp. Overall, the matched filters presented here are robust and flexible, able to accurately detect spikes from various kinds of biological data.

1.2 Executive Summary: Classification

In most cases it is not enough to simply detect a signal of interest. Many if not most studies of biological systems strive to compare different scenarios against one another, or distinguish between the actions of several forces participating in a single phenomenon; this analysis process is called classification. Classification is a two part problem. First the salient features of the signal must be extracted. Then those features have to be clustered into classes. In this work we concentrate on feature extraction, and use relatively well known algorithms for clustering.

Our first feature extraction approach was developed to classify neurons from multi-sensor extracellular neuronal recordings. The method we developed, MUSIC, is unsupervised and focuses on finding the neurons' locations from their recorded APs. These locations are then used as classification features. The second approach presented here was developed to distinguish between cardiomyocyte DEs affected by different drugs. We selected salient DE features – upslope, width, and downslope – based on known cardiomyocyte DE behavior, and compared then using a Kolmogorov-Smirnov test.

1.2.1 MUSIC

Neuron location is a classification feature that offers both stability against noise and remains invariant in time and space. Several neuronal source localization algorithms have been proposed and used *in vivo* [11, 14, 4, 94, 64], employing various approaches including simple heuristic AP generative models [14, 4], as well as monopole [11], dipole [64], and line source [94] approximations.

We explored a different localization method rooted in statistical signal processing [53]. The multiple signal classification (MUSIC) algorithm has proven effective in our preliminary

source localization experiments with tetrodes [53]. Here we tested the MUSIC algorithm, as an AP classification feature extraction method, on tetrode as well as heptode extracellular recordings.

1.2.1.1 MUSIC Algorithm Design

In the most basic case, we can treat the neuron as a point source and the surrounding medium as an isotropic, homogeneous volume conductor. Although simplistic, the monopole model has often been used in application to neural source localization [14, 4, 2, 11], and will be used here as the basis of our forward model.

Consider a static linear system that outputs a C-sensor signal generated by a single source. The system's response to a unitary signal input is the lead field vector (LFV), $m \in \mathbb{R}^{C\times 1}$. The MUSIC algorithm finds the source location $r^\star$ for which m is most orthogonal to the noise subspace [89]

$$r^\star = \arg\min_r \frac{m^T(r)E_N E_N^T m(r)}{m^T(r)m(r)} \tag{1.2}$$

where $E_N \in \mathbb{R}^{C\times(C-1)}$ is the noise subspace, calculated using singular value decomposition.

In the case of a single monopole-like source the LFV takes on the form of an electric monopole with unitary charge located at some distance, d_c from each recording sensor c. In order to enhance the feature space, we also developed an adjusted monopole model that accounts for different recording sensor impedances. This model is referred to as the impedance model.

1.2.1.2 Applications

The MUSIC algorithm for feature extraction was applied to two different data sets - one collected with a tetrode, and another collected with a heptode. The first application, using data from the tetrode, applies the MUSIC monopole model and clusters the resulting features with an expectation maximization and Bayes' information criterion (EM-BIC) based approach. Although no ground-truth was available, the resulting classified waveforms were compared, and inter- and well as within-class analyses were preformed to test classification efficacy. The second application, using data from the heptode, employs the MUSIC impedance model and then clusters the resulting features with density-based clustering of applications with noise (DBSCAN). In this case t-tests were used to compare waveforms across different classes.

Monopole Model with Tetrode

The data and results presented here are adapted from Szymanska et al. 2013 [99]. Briefly, localization was performed using MUSIC under a monopole model. Six distinct source neurons were classified from 20 seconds of extracellular tetrode recordings. On average, 89.5% of the waveforms making up each class matched the shape of the average class waveform. These results indicate that this classification scheme can successfully identify individual neurons from multi-sensor AP recordings.

Methods Twenty seconds of data from four sensors of a planar microelectrode placed in an adult locust's antennal lobe were used for this study [78, 79]. Spike detection was performed using the supervised matched filter for multi-sensor extracellular recordings presented in the previous section [102]. The source of each detected AP was then localized using the MUSIC algorithm, with a monopole source model. The EM algorithm was used to group MUSIC-derived source locations into several different cluster models. The optimal cluster model,

or number of clusters, was determined by maximizing the BIC across all models [90]. The APs from each cluster were aligned to their peak values and averaged to demonstrate the representative waveforms for each cluster. Waveform signatures, representing the relative signal power across the four sensors, were then calculated for each AP, and used to determine the within-class consistency between the waveforms.

Results All 1040 APs detected in the 20 second data stream were successfully localized using the MUSIC monopole model. This data was classified yielding 6 distinct location clusters. Only 34 source locations were classified as outliers, representing 3% of the data set. The underlying AP waveforms representing each source location in a given cluster were analyzed to assess classification efficacy.

Waveform signatures, defined here as the ranking of signal power across the four sensors for each AP, were compared within clusters. For 4 of the clusters, N1, N3, N4, and N5, the waveform signatures were consistent with the average waveform signature among 98.9%, 95.4%, 98.1%, and 90.1% of the APs in each cluster, respectively. It is reasonable to conclude that these clusters represent unique and singular neurons. The remaining two clusters, N6 and N2, were slightly less consistent, with 77.2% and 77.5% of the APs, respectively, matching their average waveform signatures. For the first one of these clusters, N6, the remaining 22.8% of waveforms match the waveform signature of a different cluster and therefore seem to be mis-classified. The last cluster, N2, contained a broad spectrum of APs in the 22.5% that did not match the average waveform. This may imply that some APs were either mis-classified or some may have been a superposition from two or more neurons. Although these clusters were less internally consistent, the results still indicate that they represent distinct neurons.

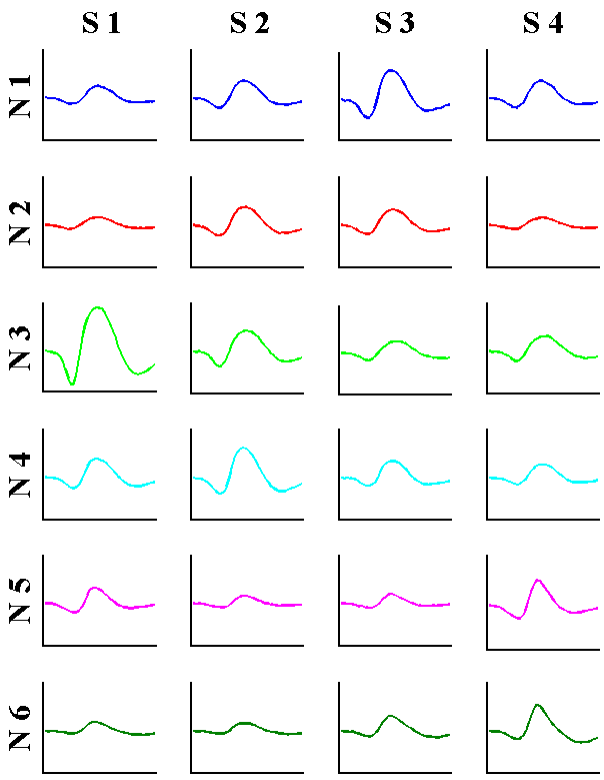

Figure 1.4: Average waveforms for each class, N1-N6, identified using MUSIC-derived source locations as features and EM-BIC for clustering. Note that each class has a unique and distinguishable waveform signature.

Conclusion Overall, classification based on our MUSIC monopole model feature extraction method successfully differentiated 6 unique and distinct classes from tetrode recorded APs, with an average accuracy of 89.5%. The average AP waveforms for each class were unique and distinguishable from each other. This is strong evidence that each class, represents a unique neuron. Given our results, this technique presents itself as a strong candidate for broad use in extracellular signal analysis.

Impedance Model with Heptode

This study demonstrates the efficacy of a multi-sensor AP classification scheme using a proxy for source "location", calculated with the MUSIC impedance model, as a classification feature. The approach was tested on extracellular heptode recordings, which we believe

will provide better resolution and stability against noise than the previously tested tetrode recordings. Five to six statistically distinct source neurons could be classified from a given heptode recording, across the 3 independent recordings tested. This shows that the method is effective for multi-sensor spike sorting.

Methods Extracellular heptode recordings were made through 1-2 mm burr holes above M1 of the right hemisphere of two male Wistar rats. A motorized headstage was used to lower the heptode into the cortex until a high activity and SNR recording site was reached ($\sim 700 - 900\mu$m into the cortex). Once the heptode was in position, 1-2 min of spontaneous neuronal activity were recorded. A total of three experiments were performed. The first two experiments, 1.1 and 1.2, were performed on a single animal at two different recording sites, one 170μm below the other, and at 2.0% and 1.5% isoflurane, respectively. The third experiment, 2.1, was performed on the second animal at 1.5% isoflurane.

Individual APs were detected from the heptode data using the supervised matched filter for multi-sensor extracellular recordings presented in the previous section [102]. The APs were then classified using the MUSIC impedance model for feature extraction, and DBSCAN for clustering [16]. The classified APs were then aligned to their peak values and averaged to demonstrate the waveform signature for each cluster. Peak AP values were compared using t-tests (95% confidence level) across each recording sensor to quantitatively assess if classes were statistically distinct.

Results 50% to 93% of the detected APs were successfully localized using the MUSIC impedance model, depending on the experiment, with overlapping and low-SNR APs accounting for most of the waveforms that could not be localized.

The MUSIC impedance model "localized" the APs to be tightly clustered around the heptode sensors. DBSCAN identified 5, 6, and 6 classes from these features for experiments 1.1, 1.2,

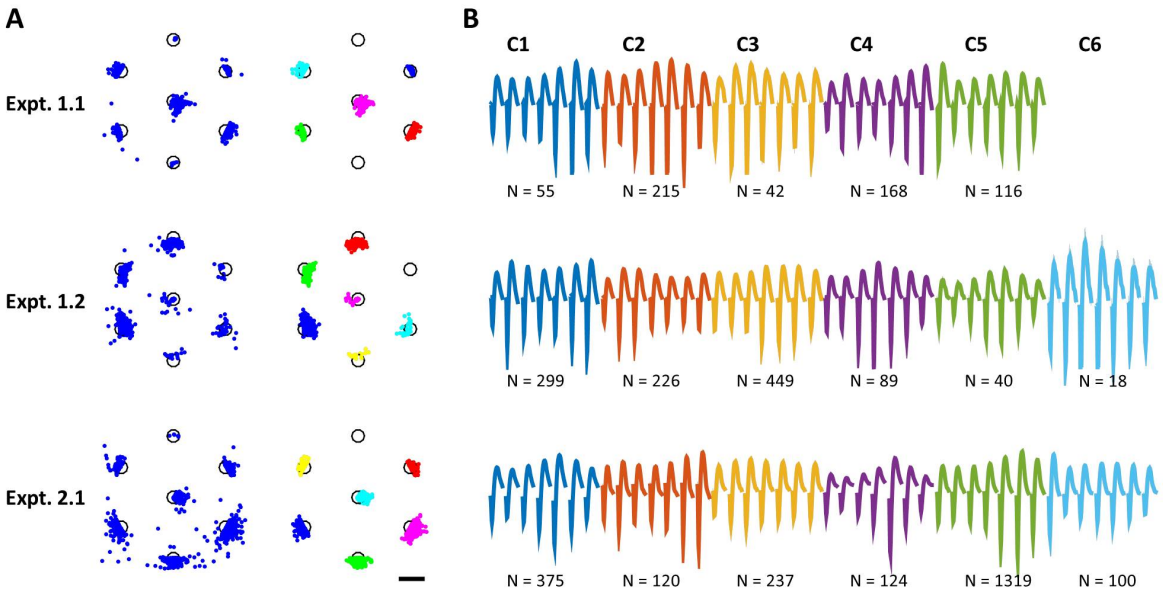

Figure 1.5: Spike sorting results using the MUSIC impedance model and DBSCAN. **A** An aerial view of the feature space and clustering results. The black bar in the bottom right corner represents 10 μm. (*Left*) Results of MUSIC localization for each experiment. The black circles represent the heptode sensors, and the blue dots are MUSIC determined source locations for each localized AP. (*Right*) DBSCAN clustering results for the features presented on the left, with outliers removed. **B** Corresponding average waveforms for each cluster identified in **A**. The number of APs in each class is listed below the waveform. T-tests revealed that the waveforms within a given experiment are statistically distinct across classes.

and 2.1 respectively. T-tests (95% confidence interval) were performed on the peak values of the waveforms for each channel to quantitatively determine if the classes within a given experiment were statistically distinguishable in the time domain. Each channel was compared to the corresponding channel for another class (CH1 Class 1 vs CH2 Class 2 etc.). In all case, at least 5 of the channels showed a statistical difference between the classes, and in most cases all 7 were statistically different. To our mind, this indicates that all of the classes in a given experiment are statistically distinguishable from each other.

Conclusion The MUSIC impedance model successfully classified 5-6 source neurons for the 3 experiments presented here. Although not all of the sources could be localized using the MUSIC impedance model, most of the feature extraction outliers were due to superimposed

APs, which were unfortunately very common in these data sets. Other APs that could not be localized suffered from low SNR. These are both clear limitations of the technique, especially compared to the base MUSIC monopole model. On the other hand, the features extracted using the MUSIC impedance model clustered more tightly than those achieved using the monopole model. The classes are also very consistent, and appear to be less prone to misclassification. Overall, the MUSIC impedance model was effective at classifying APs from extracellular heptode data. Furthermore, both qualitative analysis as well as t-test showed that the AP waveforms were consistent within each class, and that the waveforms were statistically different between classes.

1.2.2 DEC

The work presented here is adapted from Szymanska et al 2016 [103]. Briefly, this study used 2-photon microscopy of fluorescent VSDs to capture the membrane voltage of actively beating human induced pluripotent stem cell-derived cardiomyocytes (hiPS-CMs). We built a custom and freely available Matlab software to quantify, and compare DEs of hiPS-CMs treated with the β-adrenergic drugs, propranolol and isoproterenol. The classification approach aims to distinguish the drug treated cardiomyocyte APs by extracting salient AP features, such as upslope, width, and downslope, and then comparing them across drug treatments using a Kolmogorov-Smirnov (K-S) test. The software, depolarization event comparison or DEC, was able to distinguish control DEs from drug-treated DEs both immediately as well as 10 min after drug administration.

1.2.2.1 Methods

Spontaneously beating hiPS-CMs were stained with VSD, and treated with the β-adrenergic drugs propranolol and isoproterenol. The cultures were imaged immediately after addition of

drugs (less than 60 sec of exposure) and again 10-15 min after addition to ensure complete exposure. One culture was treated with propranolol, and two cultures were treated with isoproterenol. Two cultures were left untreated and imaged as controls. The imaging data was then processed to extract fluorescence intensity traces along a given cell membrane, and filtered to remove photobleaching artifacts.

Once the DEs were detected from each data trace, using the MaD detector, the individual spikes were extracted from the data and normalized such that the average DE for any given data trace had a minimum value of 0 and a maximum value of 1. This approach allowed us to preserve within data trace variations around the average DE, while also normalizing DE amplitudes across different data traces. Average waveforms were calculated for each data trace, and the full-width half-max (width), the positive slope at half-max (upslope), as well as the negative slope at half-max (downslope) were calculated for each identified DE. These parameters were compared as a function of drug treatment and time elapsed after drug treatment using a two-sample Kolmogorov-Smirnov (K-S) test at a 5% significance level.

1.2.2.2 Results

A total of 7 data traces, from 5 cell cultures, were tested. The first 5 conditions are a control (cell culture 1, Control-1), immediately after addition of isoproterenol (cell culture 2, Iso-1 0min), 10 min after addition of isoproterenol (cell culture 2, Iso-1 10min), immediately after addition of propranolol (cell culture 3, Pro-1 0min), and 10 min after addition of propranolol (cell culture 3, Pro-1 10min). In order to better test DEC in low-SNR environments we also provide results for a low-SNR control data case (cell culture 4, Control-2), as well as a low-SNR 15 min after addition of isoproterenol case (cell culture 5, Iso-2 15min).

The higher SNR (SNR $\geq$ 5) drug treated data (Iso-1 and Pro-1) were compared to the higher SNR control (Control-1). Both the Iso-1 0min and Pro-1 0min drug treatments

were statistically distinguishable from Control-1 in width, upslope, and downslope. These difference was maintained 10 min after drug administration (Iso-1 10min and Pro-1 10min). To show that the method is also applicable in a low-SNR setting, we performed a DE comparison analysis on Control-2 (SNR = 3.07) and Iso-2 15min (SNR = 0.63). The Iso-2 15min drug treatment was statistically distinguishable from Control-2 in width, but not in either upslope or downslope. This is expected, as the DE width difference should be the most pronounced and therefore least subject to noise.

The drug treatments were then compared in time to evaluate if the 0 min cases, exhibiting initial shock from the drugs, could be distinguished from the 10 min cases, which should exhibit a more stabilized response. The Iso-1 0min case was statistically distinguishable from the Iso-1 10min case in all three parameters. The Pro-1 0min case was statistically distinguishable from the Pro-1 10min case in width and upslope.

1.2.2.3 Conclusion

In summary, KS-tests of DE widths, upslopes and downslopes revealed that DEs immediately after either isoproterenol or propranolol administration are distinguishable from controls, and that these differences are maintained 10 min after drug administration. DEs immediately after drug administration were also distinct from DEs 10 min after drug administration. Lastly, isoproterenol treated cells were distinguishable from controls even at SNR ≤ 3.

These results indicate that DEC can accurately distinguish drug-treated DEs from controls, even at low SNRs, and can also distinguish drug-treated DEs based on the time after drug administration. Overall, DEC is a useful new tool for the study of cardiomyocyte electrophysiology. Combined with the use of VSDs, it allows for non-invasive, image-based, and automated analysis of cardiac DEs. This study demonstrates the ability of this tool to quantify changes in DEs as a function of drug treatment and as a function of time.

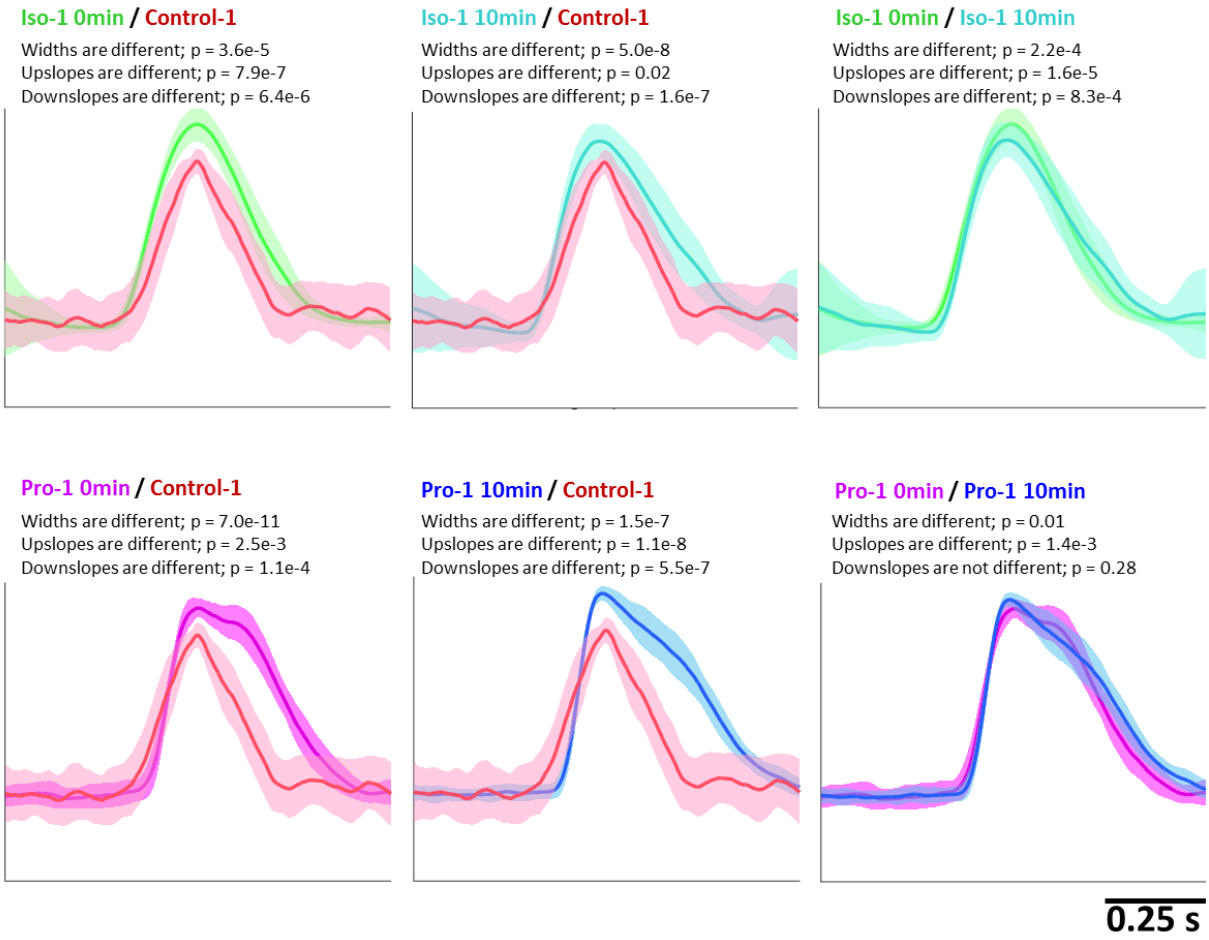

Figure 1.6: DE analysis with respect to drug treatment and time after drug administration. 6 panels are shown. Each panel shows the average DEs of the two specified drug treatments, with standard deviations shaded around the average. The title of each panel identifies the two treatments being compared. The text below the title indicates whether the DE populations were different in width, upslope, and downslope, and provides p-values. The top 3 panels, going from left to right, compare Iso-1 0min (green) to Control-1 (red), Iso-1 10min (light teal) to Control-1, and Iso-1 0min to Iso-1 10min. Both Iso-1 0min and Iso-1 10min were different from Control-1, and were different from each other, in all three parameters (width, upslope, downslope). The bottom 3 panels, going from left to right, compare Pro-1 0min (purple) to Control-1 (red), the Pro-1 10min (blue) to Control-1, and Pro-1 0min to Pro-1 10min. Both Pro-1 0min and Pro-1 10min were different from Control-1 in width, upslope, and downslope. They were also different from each other in width, and upslope.

1.2.3 Conclusion

In this section we presented two approaches to electrophysiological data classification.

The first classification problem focused on how to distinguish an unknown number of neurons from multi-sensor extracellular recordings of neuronal APs. AP sources were localized using MUSIC, as a feature extraction step, and the approach was tested on both tetrode and heptode data. The algorithm applied to the tetrode data used the MUSIC monopole model for feature extraction, and EM-BIC for clustering. Six distinct clusters were identified, all exhibiting different AP waveform signatures. On average the within-class classification consistency was 90%. Overall, this shows that the MUSIC monopole model used with an EM-BIC clustering tool is an accurate classification method that can distinguish several neurons from tetrode data. The algorithm applied to the heptode data used the MUSIC impedance model for feature extraction, and DBSCAN for clustering. The resulting clusters were very tightly grouped in the feature space. The average AP waveforms for each class were statistically different, as determined by t-tests, for all tested data sets. Furthermore, the classes were very consistent, implying a low rate of misclassification. Overall, the MUSIC impedance model used with DBSCAN proved to be an accurate tool for sorting spikes from heptode data.

The second classification problem focused on how to distinguish the DEs of actively beating VSD imaged caridomyocytes, treated with 2 different β-adrenergic drugs (propranolol and isoproterenol). Our algorithm, DEC, compared salient DE features - upslope, width, and downslope - across the two different drug administrations using a K-S test. The drug treated cells' DEs were statistically different from the controls both immediately after as well as 10-15 min after drug administration for both drugs, even at SNR ≤ 3. DEs immediately after drug administration were also distinct from DEs 10 min after drug administration. Overall,

DEC's ability to accurately distinguish drug treated DEs from controls, even under a low-SNR environment, makes it a valuable tool for cardiomyocyte drug assay studies.

1.3 Executive Summary: Discovery

The ultimate goal of developing the signal processing techniques outlined in the previous sections is to apply them to electrophysiological data in the process of discovery. This next section outlines exactly this process. Both the multi-sensor extracellular AP detector as well as the MUSIC-based classification scheme came together with a few other statistical methods to analyze extracellular heptode data looking at neuronal circuit dynamics as a function of anesthesia depth.

The study presented in this chapter is adapted from Szymanska et al. 2017 [101] and takes a multi-modal approach. Three types of neurological recordings were made simultaneously - electrocorticography (ECoG), multi-unit activity (MUA), and local field potentials (LFP) - to investigate how regional signals from the brain surface relate to neuronal activity from deep cortical layers during varying levels of anesthesia. We found that higher anesthesia led to higher AP frequency, no change in the number of active single units, and increased cross-hemisphere functional connectivity. Additionally, all APs were restricted to ECoG bursts, with no APs occurring during suppressed ECoG states.

1.3.1 Methods

Two male Wistar rats were implanted with two ECoG screw electrodes for recording over M1 and V1 of the left hemisphere. Extracellular heptode recordings were made through a 1-2 mm burr hole above M1 of the right hemisphere. A motorized headstage was used to lower the heptode into the cortex until a high activity and SNR recording site was reached

($\sim 700 - 900\mu$m into the cortex). Once the heptode was in position, 1-2 min of spontaneous neuronal activity were recorded at three monotonically increasing isoflurane anesthesia levels - low, intermediate, and deep.

A total of three experiments were performed. The first two experiments, 1.1 and 1.2, were performed on a single animal at two different recording sites. The third experiment, 2.1, was performed on the second animal. Each recording will be referred to by the experiment number followed by "Iso" and then the percentage of administered isoflurane.

Individual APs were detected from the heptode data using the supervised matched filter for multi-sensor extracellular recordings presented in the previous section [102]. The APs were then classified using the MUSIC impedance model for feature extraction, and DBSCAN for clustering [16]. If there were not enough APs detected ($N \leq 100$) to reliably cluster using the mentioned algorithms, the APs were classified manually. Burst suppression ratios (BSR), AP firing frequencies, and spike triggered averages (STA) were calculated for all data sets.

1.3.2 Results

The results focus on AP firing frequencies, STAs, and the number of neuron classes, as a function of anesthesia depth.

The overall spiking activity decreased as the level of administered isoflurane increased, in each of the three experiments. Interestingly, APs were only detected during ECoG bursting activity (total of $\sim$ 15 min of recording, and 12,675 detected APs). Therefore the decrease in overall spiking activity followed from the burst suppression ratio increasing. Furthermore, the average AP firing frequency within a given ECoG burst was increasing as the level of isoflurane was increasing. For experiment 1.1 the AP firing frequency began at 73 $\pm$ 18 APs/s, increased to 135 $\pm$ 47 APs/s, and then increased again to 243 $\pm$ 65 APs/s. The

same trend was observed for experiment 1.2 where the firing frequency progressed from 108 ± 32 APs/s, to 193 ± 30 APs/s, to 224 ± 31 APs/s. A similar increase, although more modest, was also observed in experiment 2.1 (125 ± 17 APs/s, to 140 ± 15 APs/s, to 155 ± 27 APs/s).These findings suggest that although overall activity decreases as the level of administered isoflurane is increased, when APs do fire, they fire at an increased frequency as the depth of anesthesia increases.

Spike triggered average analysis was performed on the data in order to investigate the relationship between the MUA and ECoG. For all three experiments and across both ECoG channels, the ECoG-STA amplitude as well as root-mean-square increases significantly as the level of isoflurane is increased. This implies a stronger degree of coupling between the left hemisphere (represented by the EEG signals), and M1 of the right hemisphere (represented by the MUA), as the level of anesthesia is increased. LFP-STA analysis showed the same results.

Spike sorting across the three anesthesia levels revealed that the same sub-population of single units was active as anesthesia deepened. For experiment 1.1, the same 5 units were identified for Iso 2.0% and Iso 2.5%. Four of the same units were also identified at Iso 3.0%. Similarly for experiment 2.1, the same 6 units were identified for Iso 1.5% 1 and Iso 1.5% 2. Five of these units were also identified for Iso 2.5%, alongside an additional unit previously not present. For experiment 1.2 a total of 9 units were identified across the three isoflurane levels. Although no single unit persisted through all three levels of isoflurane, 8 of the 9 units recurred at two of the isoflurane levels. Overall these results indicate that the same sub-population of neurons is active as levels of isoflurane are increased during burst suppression. Although a few units disappeared, and a few new ones appeared as the level of anesthesia changed, for the most part the same units could be reconciled across the different levels of isoflurane.

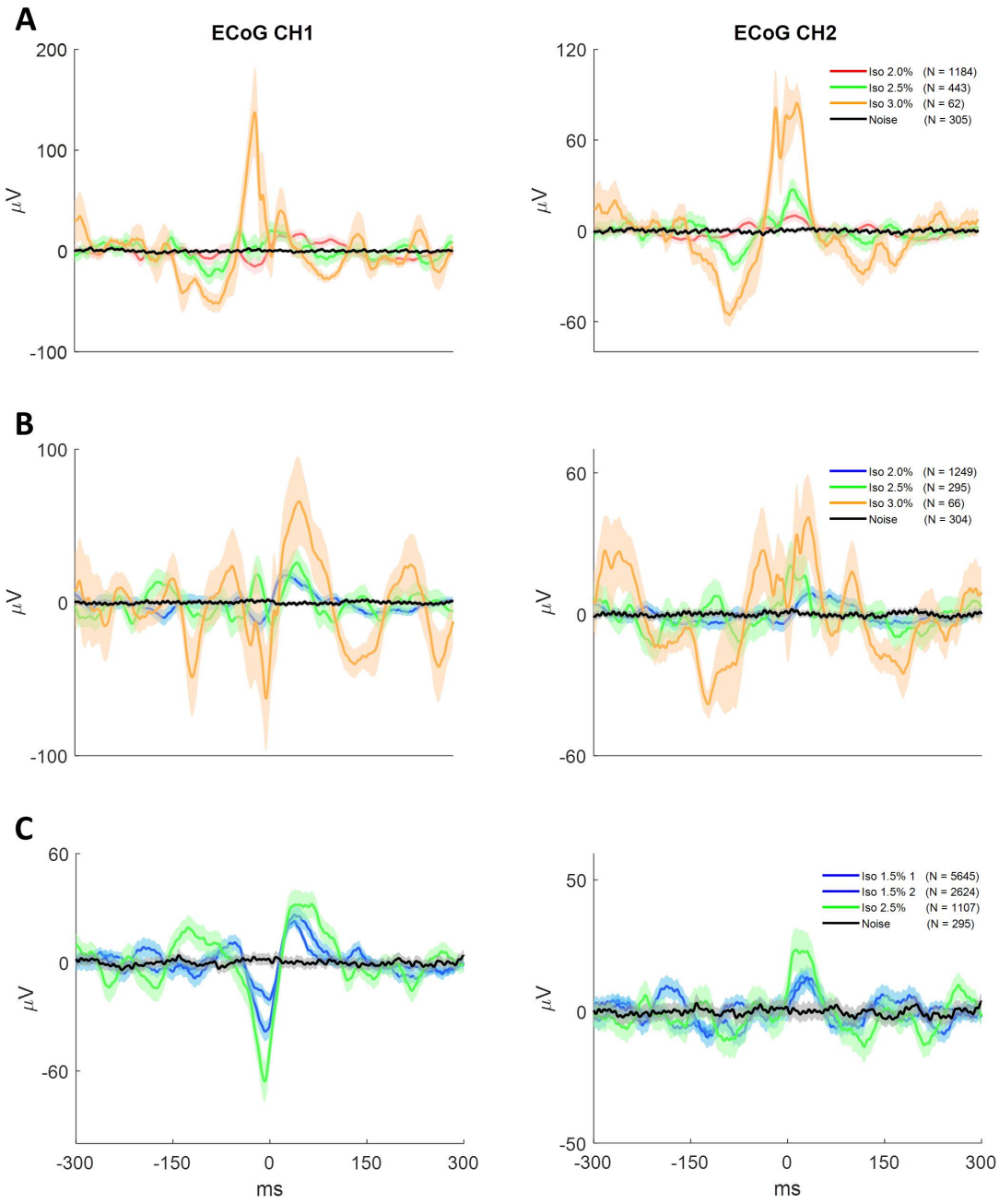

Figure 1.7: As the level of isoflurane is increased, the resulting cross-hemisphere ECoG-STA increases in magnitude for all three experiments. (**A**) (**B**), and (**C**) show ECoG-STAs at varying isoflurane levels for experiment 1.1, 1.2, and 2.1, respectively. The left column shows results for ECoG channel 1, and the right column for ECoG channel 2. The AP was centered at 0 ms for all ECoG-STAs. Shaded regions represent 95% confidence intervals and the number of APs being averaged over is depicted in the three legends. The ECoG-STAs are color coded for isoflurane level, with blue being 1.5%, red 2.0%, green 2.5%, and orange 3.0%.

1.3.3 Conclusion

Overall, our findings suggest that as isoflurane anesthesia is increased during burst suppression, neuronal activity remains bound to the same sub-population of neurons; these neurons' overall activity decreases as the burst suppression ratio goes up. However, when activity does occur it is at a higher firing frequency as the level of anesthesia is increased. Likewise, the activity is increasingly more synchronous with cross-hemisphere activity, as measured by ECoG, and local population activity as measured by LFP.

Chapter 2

Detection

Detection is a crucial component of any biophysical signal processing algorithm. Although spike detection in general has been an issue for many years, the problem is not completely resolved. Electrophysiological data is generally very noisy, therefore signal detection tools from other fields don't readily apply. The noise coming from the recording hardware and environment is in most cases white Gaussian noise and is relatively easy to process. However, the presence of biological noise is much more problematic to spike detection because it is often both correlated and statistically similar to the signal. This is especially an issue in neuronal multi-sensor extracellular recordings where the activity of background neurons and ion-channel noise contribute to a very low signal-to-ratio (SNR). Although biological noise is less prominent in many imaging recording modalities, the low SNR issue still persist, and other factors such as motion artifact correction, and photobleaching compound the detection problem.

With that in mind we have developed several detection algorithms, all based on data driven machine-learning techniques and a matched filter, to detect spikes from 1) extracellular recordings of neurons, 2) calcium imaging of neurons, and 3) voltage sensitive dye imaging

of cardiomyocytes - all electrically active cells that exhibit a "spike" as the signal of interest. The following chapter details the broad detector design, and then presents the three applications listed above. Other existing techniques, the motivation behind these applications, as well details on the detector's implementation are presented with the results for each modality. The chapter ends with some preliminary work on unsupervised approaches as well as other ideas for further automation.

2.1 Background

The most popular detection algorithms can be separated into three broad categories, raw data thresholding, abstract mathematical features, and template matching. Thresholding is the most commonly used detection method due to its computational simplicity. It is also therefore the standard that most other techniques are compared against. Overall, thresholding is a very simple and useful approach. However, the average sensitivity is about 90%, and the technique suffers tremendously in low-SNR environments, which largely limits its scope [83, 57, 7, 48, 97]. Abstract mathematical features are also a very attractive approach to spike detection. The most popular abstract mathematical detection approaches are independent component analysis (ICA), principle component analysis (PCA), and, more recently, morphological filters. Overall, abstract techniques perform better then thresholding, with an average sensitivity of about 95% [57, 107, 41]. However, the approach is again sensitive to low SNR.

2.2 Matched Filter

The matched filter is arguably the best detection approach available for deterministic signals in random noise. Although biological, and especially neurological noise is not random,

matched filtering has proven to effectively detect spikes from various types of neurological [107, 48, 122, 9, 99, 104, 47], as well as cardiological data [50, 112, 103]. The basis behind template matching is to generate a template representative of electrophysiological activity, and then compare the data to your template. Traditional template approaches are supervised, and the template is generated from APs or spikes selected by an analyst from the data [99, 103, 104, 26]. However, unsupervised algorithms using wavelets as templates have been growing in popularity [122, 91, 9, 71, 100]. In general, template matching outperforms other techniques [48, 107, 100, 104] and was therefore pursued as the detector of choice in these studies. More specifically, a supervised template approach was adopted because it outperforms wavelet approaches in the kinds of low-SNR environments often seen in electrophysiological recordings [122, 91, 48, 107, 104].

2.2.1 Matched Filter Design

Although the matched filters implemented throughout this work have analyzed very different kinds of data, including 1) multi-channel extracellular recording of neurons, 2) voltage sensitive dye imaging of cardiomyocytes, and 3) calcium imaging of neurons, they all share a common base design, detailed below. The main differences lie in parameter and implementation details, explained further in the application section (Sec. 2.3).

2.2.1.1 General Likelihood Ratio Test

According to the Neyman-Pearson lemma, the likelihood ratio test (LRT) is the most powerful discriminant of two underlying models from a deterministic signal. In the case of electrophysiological signals, the two models being investigated represent noise-only data, and data containing both spikes and noise. More formally, spike detection can be interpreted as a hypothesis testing problem, where under the null hypothesis, H_0, the signal contains noise

only, and under the alternative hypothesis, H_1, the signal contains both a spike and noise. Given a C sensor array signal of length N, where C is the number of sensors and N is the number of samples spanned by a spike, we can express the hypotheses mathematically as

$$\begin{aligned} H_0 : &\quad \underline{\mathrm{x}} = n \\ H_1 : &\quad \underline{\mathrm{x}} = \underline{\mathrm{s}} + n \end{aligned}$$

where $\underline{\mathrm{x}} \in \mathbb{R}^{C \times N}$ is a signal of length N, $\underline{\mathrm{s}} \in \mathbb{R}^{C \times N}$ is a spike, and $n \in \mathbb{R}^{C \times N}$ is zero-mean noise. The LRT for this problem can now be expressed as

$$L(x) = \tfrac{p(x|H_1)}{p(x|H_0)} \quad , \quad \begin{array}{llll} H_1 & \text{if} & L(x) > \gamma_o \\ H_0 & \text{if} & L(x) < \gamma_o \end{array}$$

where $x \in \mathbb{R}^{1 \times CN}$ is the row vector form of the signal matrix $\underline{\mathrm{x}}$, and γ_o is the threshold.

2.2.1.2 General Matched Filter

Assuming Gaussian noise statistics, the LRT takes on the form of a generalized matched filter (GMF) [46]

$$S(x) = s\Sigma^{-1}x^T \quad , \quad \begin{array}{llll} H_1 & \text{if} & S(x) > \gamma \\ H_0 & \text{if} & S(x) < \gamma \end{array} \tag{2.1}$$

where $s \in \mathbb{R}^{1 \times CN}$ is the row vector form of the spike $\underline{\mathrm{s}}$, $\Sigma \in \mathbb{R}^{CN \times CN}$ is the spatio-temporal noise covariance matrix of the row vector form of the noise n, and γ, which subsumes γ_o and data-independent terms, is the threshold. Note that the test statistic, $S(x)$, is linearly dependent on the signal x. Even if the noise is not Gaussian, $S(x)$ has the highest achievable SNR of all other linear combinations of the data [46].

2.2.1.3 Sliding Window Approach (Convolution)

The matched filter described in Sec. 2.2.1.2 performs a hypothesis test on $\underline{x} \in \mathbb{R}^{C \times N}$ where C is the number of sensors and N is defined as the number of samples spanned by a typical spike. To detect spikes from a full data trace, $X \in \mathbb{R}^{C \times T}$, where $T \gg N$, we have to apply the matched filter iteratively across the entire trace. This type of approach is referred to as a sliding window. Initially the filter tests a segment of the signal $[X(1), X(2), \cdots, X(N)]$. The filter output given this window is a scalar that indicates the likelihood of a spike in the center of the window, $S(\frac{N}{2})$. Once finished, the test segment is advanced by one sample to $[X(2), X(3), \cdots, X(N+1)]$ and this new segment is filtered. The process is repeated until the full data set X has been filtered. This is also referred to as a convolution.

2.2.2 Simplifying Assumptions

The $CN \times CN$ parameters of the covariance matrix Σ may be difficult to estimate reliably given a finite amount of data available for training. If necessary, we can reduce the number of parameters by making some simplifying assumptions about the noise statistics used to calculate the covariance matrix, Σ.

2.2.2.1 Spatially White Noise Assumption

We can reduce the number of parameters calculated in the noise covariance, Σ, by making the simplifying assumption that the noise statistics are spatially white, meaning that the noise over a given channel is independent of the noise on other channels (uncorrelated in space). To then perform the noise covariance calculation, using this simplifying assumption,

we can represent Σ as

$$\Sigma = \begin{bmatrix} \underline{\Sigma_{1,1}} & \underline{\Sigma_{1,2}} & \cdots & \underline{\Sigma_{1,C}} \\ \underline{\Sigma_{2,1}} & \underline{\Sigma_{2,2}} & \cdots & \underline{\Sigma_{2,C}} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\Sigma_{C,1}} & \underline{\Sigma_{C,2}} & \cdots & \underline{\Sigma_{C,C}} \end{bmatrix} \tag{2.2}$$

where the submatrix $\underline{\Sigma_{i,i}} \in \mathbb{R}^{N \times N}$ is the temporal covariance matrix of noise at sensor i, and $\underline{\Sigma_{i,j}} \in \mathbb{R}^{N \times N}$ $(j \neq i)$ is the temporal cross-covariance matrix of noise at sensors i and j. If the noise statistics are uncorrelated in space, then all $\underline{\Sigma_{i,j}}$ $(j \neq i)$ are 0. Although electrophysiological noise exhibits spatial correlations [100], our experience shows that the spatially white noise assumption often outperforms fully colored noise in some detection cases, namely multi-sensor extracellular APs [28].

2.2.2.2 White Gaussian Noise Assumption

Similarly, due to available noise sample constraints as well as processing time considerations, it may be necessary to reduce the number of parameters in Σ even more drastically. We can do this by assuming that the noise statistics are spatially and temporally white, meaning each noise sample is independent of any other noise sample. Again, the noise in electrophysiological data is generally not white, as a large portion of it is biological noise. However, we've seen that it is often better to use a white Gaussian noise (WGN) approximation than risking a poorly estimated full noise covariance model [103].

Under the white Gaussian noise (WGN) assumption all parameters, $\Sigma_{i,j}$ $(j \neq i)$, of the covariance matrix Σ, are zero. The remaining parameters $\Sigma_{i,i}$ represent each sample point's autocorrelation, which is equivalent to the noise variance, σ^2. The covariance matrix then takes on the form

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_{N\times N}$$

where σ is the only parameter being estimated from the noise training sample. The generalized matched filter is then reduced to a matched template

$$S(x) = s\frac{1}{\sigma^2}x^T \quad , \quad \begin{array}{lll} H_1 & \text{if} & S(x) > \gamma \\ H_0 & \text{if} & S(x) < \gamma \end{array} \tag{2.3}$$

2.2.3 Setting Detection Thresholds

Thresholding parameters were calculated from the filtered signal which contains both spikes and noise. Assuming that spikes are statistically sparse within the noise, we can effectively approximate the mean and standard deviation of the filtered noise for each data trace X by calculating the filter output's median, M, and median-based standard deviation, σ_M [71].

$$\sigma_M = \frac{\mathcal{M}\{\mid S(1)) - M \mid, ..., \mid S(T) - M \mid\}}{\sqrt{2}Erf^{-1}(\frac{1}{2})} \tag{2.4}$$

where S is the filter output, $\mathcal{M}$ is the median operator, and T is the number of time samples spanning X. This median-based approximation provides a more robust estimate of the filter output's noise parameters than simply taking the mean and standard deviation of the filter output. The threshold, γ is then set as a multiple of σ_M, above the median, M

$$\gamma = M + a\sigma_M \tag{2.5}$$

where a is the multiple and will be referred to as the thresholding sensitivity for the remainder of this work.

2.3 Applications

The matched filter is data driven and therefore very flexible by design. Throughout this work it's been specifically tailored to three applications, on three very disparate data types. The first is multi-sensor extracellular recordings, where the purpose is to detect neuronal action potentials (APs). The second application strays from electrical recordings and moves into imaging; the matched filter was applied to detect depolarization events (DEs), which are often also called APs, from voltage sensitive dye (VSD) imaged cardiomyocytes. The third and last application remains in the realm of imaging, and goes back to neurological data; the filter was applied to detect calcium events (CEs), a proxy for AP activity, from calcium dye imaged neuronal cell populations as well as dendritic spines.

2.3.1 Multi-sensor Extracellular Action Potential Detection

The work presented in this section is adapted from Szymanska et al. 2013 [99]. Briefly, a matched filter was designed and implemented to detect extracellular APs from multi-sensor extracellular recordings. The detector was tested on tetrode data from a locust antennal lobe and assessed against three trained analysts. 25 APs and noise samples were selected manually from the data and used for training. To reduce complexity, the filter assumed that the underlying noise in the data was spatially white. The detector performed with average TP and FP rates of 84.62% and 16.63% respectively. This high level of performance indicates the algorithm is suitable for widespread use.

2.3.1.1 Background on Extracellular Action Potential Detection

Before any signal analysis can be performed, APs representative of neurophysiological activity must be identified from the data. Detection is therefore a crucial component of any

neurophysiological signal processing algorithm. Biological noise in extracellular neurophysiological recordings is composed of the activity of background neurons and ion-channel noise. It is therefore both correlated with, and statistically similar to the AP signal. This compounds the detection problem and makes most standard signal detection tools used in other fields unsuitable.

Template matching, or more generally the matched filter, have proven to effectively detect spikes from various types of neurophysiological data [107, 48, 99, 122, 9, 26, 100, 91, 71]. Although widely applied to single-sensor extracellular data, there are few examples of matched filers applied to multi-sensor data. Gozani and Miller [26] developed a technique using multiple matched filters to simultaneously detect and classify multi-sensor APs. Their technique tries to both maximize AP signal-to-noise ratio (SNR) and minimize the interference between APs. However, this method relies on the construction of many AP templates, which requires some prior knowledge of the number of recorded neurons, and may be very time consuming, as a single tetrode in a neuron-dense region may record activity from up to 20 neurons [115]. Furthermore, minimizing the interference between APs may compromise their detectability. The filter presented here concentrates on a more reliable and simpler supervised method that uses only one template and takes into account the noise statistics of the data.

2.3.1.2 Methods

Data Collection Data used in this experiment is publicly available online [78] and the full data collection procedure is described in Pouzat et al. 2002 [79]. Briefly, a planar silicon probe was placed below the surface ($\sim$50-100 μm) of an adult locust's antennal lobe and used for recording. Recorded data was sampled at 15 kHz and bandpass filtered from 300 - 5,000 Hz. A total of 20 seconds of data was provided from four of the probe tip sensors. Half of the collected data was used for training (training data), and the remaining half was

used for further analysis (test data). Three trained analysts independently tagged all of the spikes in the test data.

Filter Training Twenty five APs, 2 ms each ($N = 30$, $C = 4$), were manually selected from the training data, aligned to their peak values, and averaged to generate the matched template, s. Similarly, 25 noise samples, roughly 30 ms each, were manually selected from the training data, concatenated into a single four-sensor time series, and used to generate the noise covariance matrix, Σ. From experience, we selected the optimal noise parameter estimation model for this type of data as spatially white noise (Sec. 2.2.2.1).

The training noise was subdivided into 375 noise windows ($N = 30$) and auto-covariance sequences, $r_c(k)$, were then calculated at lags $k \in \{-29, \cdots, 29\}$ for each window and each sensor $c \in [1, 4]$. The sequences were averaged across all 375 noise windows, and used to generate each $\underline{\Sigma}_{i,i}$ (Eq. 2.2), which were then used to construct Σ.

Performance Analysis Detected APs, given 40 incrementally increasing thresholds, $a = [1, 1.5, ..., 20.5]$ (Eq. 2.5), were compared against those tagged by each trained analyst, with the analyst acting as the "ground-truth". True positive (TP), and false positive (FP) rates for each threshold and analyst were then calculated as

$$\text{TP Rate} = \frac{TP}{Spikes_A} \tag{2.6}$$

$$\text{FP Rate} = \frac{FP}{Spikes_D} \tag{2.7}$$

where $Spikes_A$ is the total number of spikes tagged by the analyst, and $Spikes_D$ is the total number of spikes detected by the detector, and TP and FP are the total numbers of true

positives, and false positives, respectively. The false negative (FN) rate can be calculated as $(1 - FP)$, however as it gives no new information about the detector performance, it is not reported here. The TP and FP rates at each threshold were then used to generate receiver operating characteristic (ROC) curves for each analyst.

2.3.1.3 Results

The detector performed well compared with all three analysts. The optimal threshold for each analyst (Table 2.1) was determined by minimizing the distance between their respective ROC curve (Fig. 2.4) and theoretically perfect performance (100% TP, 0% FP). At the optimal threshold, the detector performed with a TP rate of 90.79% and an FP rate of 20.66% compared with Analyst 1, a TP rate of 80.00% and an FP rate of 10.31% compared with Analyst 2, and a TP rate of 83.06% and an FP rate of 18.91% compared with Analyst 3 (Table 2.1). The detector's average TP and FP rates were 84.62% and 16.63%, respectively.

Note that the optimal thresholds associated with the best performance for each analyst differed. This is a result of the discrepancy in spike tagging between analysts, which can be quantified in terms of signal-to-noise ratio (SNR). The SNR for each spike, u, tagged by the analysts was calculated as

$$\mathrm{SNR}(u) = \max_{c} \left\{ \frac{\sigma_c^2(u)}{\sigma_c^2(n)} \right\} \tag{2.8}$$

where $\sigma_c^2(u)$ is the variance of spike u at sensor c, and $\sigma_c^2(n)$ is the variance of the training noise at sensor c. The maximum SNR across sensors was taken as u's SNR. Analyst 1 was the most selective with a median SNR of 28, followed by Analyst 2 with a median SNR of 20, and Analyst 3 was the most liberal, with a median SNR of 7. An example of the types of spikes tagged by each analyst, as well as the detected spikes at $a = 5.5$ is shown in Fig. 2.2.

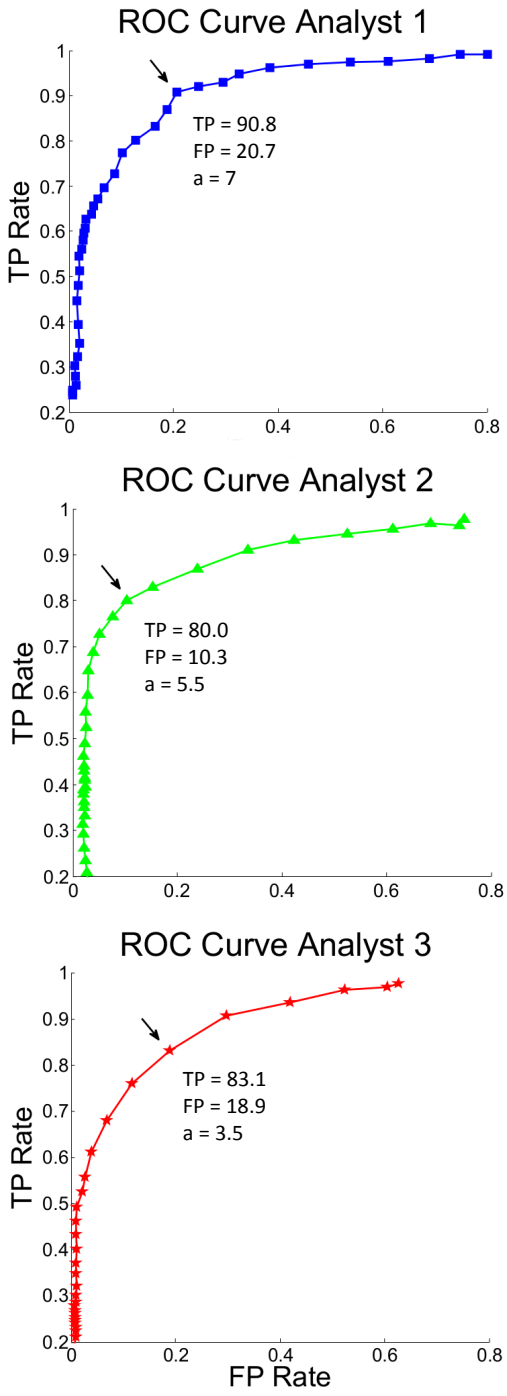

Figure 2.1: ROC Curves for all three analysts. TP and FP pairs were collected for 40 incrementally increasing thresholding sensitivity with $a = [1.0, 1.5, ..., 20.5]$ (Eq. 2.5), for each analyst. The black arrows point to the detector's optimal performance for each analyst, with the appropriate parameters listed below.

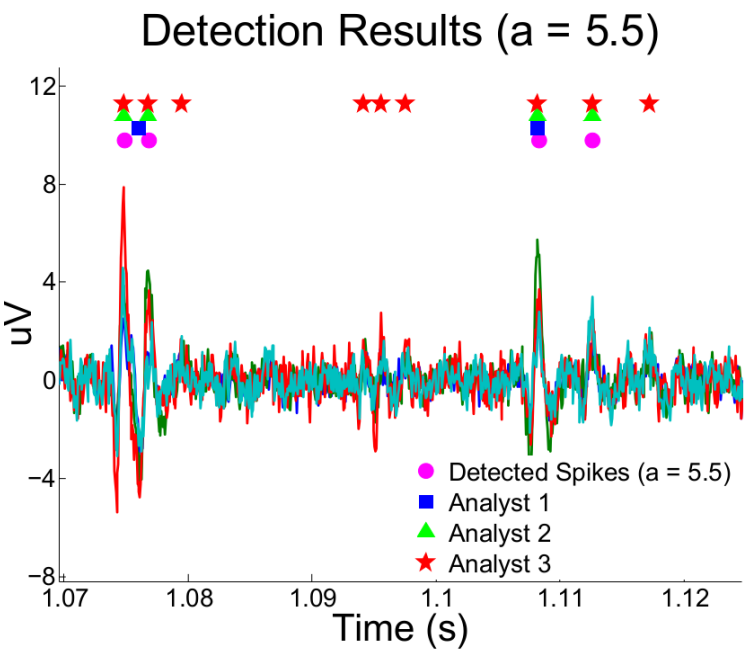

Figure 2.2: Raw tetrode data with the detection results for $a = 5.5$ marked as purple circles. The spikes tagged by Analysts 1, 2, and 3 are shown as blue squares, green triangles, and red stars, respectively. Note that this threshold is optimal for detecting the spikes tagged by Analyst 2.

Table 2.1: Optimal detector performance for the three analysts. The corresponding thresholding sensitivity a (Eq. 2.5) is listed in column 3 and each analyst's median spike SNR is provided for reference in column 4.

Analyst	Optimal Performance		Threshold	Median
	TP Rate (%)	FP Rate (%)	Multiple 'a'	Analyst SNR
1	90.79	20.66	7	28
2	80.00	10.31	5.5	20
3	83.06	18.91	3.5	7

2.3.1.4 Discussion

Analyst Reliability The data used in this study was collected *in vivo*, therefore the ground-truth in terms of spike times was not available and was instead estimated as the spikes selected by each analyst. However, agreement between analysts was as low as 39%. The variability in the analysts' tagged spikes therefore implies that human analysts may not be a reliable source of ground-truth estimation.

Besides not being consistent with each other, the analysts were also not internally consistent in spike selection. Analysts were likely to arbitrarily increase or decrease their internal thresholding criteria over time, even while tagging spikes in a single data set. Similarly, spikes with a relatively low amplitude but surrounded by noise were very likely to be selected by the analyst. However, similar spikes surrounded by other APs were less likely to be selected. This bias artificially decreases the detector's TP rate. In general human analysts are not a reliable or consistent source of ground-truth for the detection of neurophysiological activity. This conclusion was also drawn by Harris et al. [32] where 9 analysts sorting extracellular tetrode data had FN and FP rates as high as 30%.

Detector Performance The presented algorithm's performance is under estimated in this study. As discussed previously, this is partially due to the analysts' unreliability. To try to better assess the detector's performance, the FPs in the optimal cases for Analyst 1 were further inspected. The detector performed optimally when compared against Analyst 1 at $a = 7$. In this case the detector reported a 20.66% FP rate (Table 2.1). Because Analyst 1 is the most selective, we can compare these FPs to the spikes tagged by both Analysts 2 and 3. The comparison results show that 99% of these FPs were tagged by both Analysts 2 and 3. In other words, a majority of the analysts thought that 99% of the spikes identified as FPs in this case were actually TPs. This implies that the FP rate should be closer to 0%. Similar results were derived when analyzing the FPs identified against Analyst 2.

Given these results, we believe that the detector would perform much better if compared against the actual ground-truth as opposed to trained analysts. This can be accomplished by patch-clamping several neurons and simultaneously recording extracellularly with a tetrode as described in [32]. The patch-clamp recordings would be temporally matched against tetrode recordings to determine precise AP times. These could then be used as the ground-truth for detection results. Any spikes recorded from non-patched neurons would not be included in such an analysis.

2.3.1.5 Conclusion

The multi-sensor matched filter developed in this study was assessed against three trained analysts and performed with average TP and FP rates of 84.62% and 16.63%, respectively. The detector's performance presents it as a great candidate for wide-spread use as one of the only multi-sensor detectors of its kind. Furthermore, the algorithm's performance is likely under estimated as a vast majority of the spikes identified as FPs for Analysts 1 and 2 were actually identified as TPs by Analyst 3. Because the analysts were proven to be inconsistent and unreliable, we believe that the detector would perform even better if compared against the ground-truth instead of trained analysts.

2.3.2 Matched-filter for Depolarization Event Detection (MaD)

The work presented here is adapted from Szymanska et al. 2016 [103]. Briefly, we used 2-photon microscopy of fluorescent voltage-sensitive dyes (VSDs) to capture the membrane voltage of actively beating human induced pluripotent stem cell-derived cardiomyocytes (hiPS-CMs). We built a custom and freely available matched filter for depolarization event detection (MaD) of VSD imaged hiPS-CMs. The efficacy of our software was quantified by comparing detection results against manual DE detection by expert analysts. The soft-

ware accurately detected DEs with true positive rates of 98-100% and false positive rates of 1-2%, at signal-to-noise ratios (SNRs) of 5 and above, given three hiPS-CM drug treatments: propranolol, isoproterenol, and control. MaD was part of a software package called MaDEC which also quantified and compared the DEs treated with these β-adrenergic drugs (Chapter 3.2).

2.3.2.1 Background on VSD Spike Detection

The derivation of human induced pluripotent stem cells (hiPS) from somatic human cells has opened broad opportunities in the study of human cardiac cells. Previously limited by their minimal proliferation, human cardiomyocytes were difficult to obtain in significant number to allow widespread study. hiPS cells can be expanded into the required quantities and then differentiated into cardiomyocytes (hiPS-CM) as a new, and seemingly endless, source of cardiomyocytes [65, 123]. Accompanying this expanded availability, there has been an acceleration of the development of new methods for assessing the electrophysiological effects of drug compounds. Image-based tools for assessing excitable cells, such as cardiomyocytes, have particularly come to the fore [8, 20, 50, 97, 109, 112], with voltage-sensitive dye (VSD) imaging proving to be particularly popular [34, 120, 24, 33, 55].

Quantitative assessment of the heterogeneity of the DEs of actively beating cardiomyocytes is critical to understanding the electrophysiology underlying these cardiac tissue models. However, existing quantitative analysis tools studying cardiomyocyte activity were developed for calcium imaging applications [97, 50, 8, 3], and most are focused on automated region of interest (ROI) identifications [97, 8, 3]. These image analysis tools are not readily applicable to VSD imaging of cell membranes, where two-dimensional regions of interest, are not necessarily spherical and are more difficult to identify. Algorithms focusing on line-scan data have been developed for Ca^{2+} spark detection in cardiac ventricular myocytes [50], as

well as elementary calcium release events in muscle [112]. However, they have not been applied or tested with VSD based methods on cardiomyocyte DEs.

To quantitatively assess VSD acquired imaging data from hiPS-CMs, we built a custom Matlab software capable of detecting individual VSD imaged DEs, using a generalized matched filter, providing non-biased as well as reliable event selection even in low-SNR environments.

2.3.2.2 Methods

Data Collection The full data collection procedure is described in Appendix A.1 as well as Szymanska et al. 2016 [103] and Heylman et al. 2015 [34]. Briefly, prepared hiPS-CMs began spontaneously beating on approximately Days 12-15 of culture, and were stained with VSD and imaged on Day 33. After staining with VSD, cells were qualitatively confirmed to still be spontaneously beating after which β-adrenergic drugs were added to the cultures. One culture was treated with propranolol, two cultures were treated with isoproterenol, and two cultures were left untreated as controls. Two-photon microscopy line scan data was then collected for all of the cultures. The drug treated cultures were imaged immediately after addition of drugs (less than 60 sec of exposure) and again 10 min or 15 min after addition to ensure complete exposure. This accounts for a total of 7 data traces, from 5 cell cultures. The imaging data was then processed to extract fluorescence intensity traces along a given cell membrane, and filtered to remove photobleaching artifacts. Intensity traces, X, for all drug and control conditions were then plotted in Matlab, and three trained human analysts independently identified DE peak times from each trace.

Filter Training The MaD detector is completely data driven as both s and Σ from Eq. 2.1 are estimated from the data. This allows the detector to be very flexible in accommodating various DE sizes, shapes, and durations, depending on the drug treatment applied and the

specific data collected. The detector was trained under three detection conditions. The first was no drug exposure (control); the second was isoproterenol exposure and included data immediately after and 10 min after addition of isoproterenol; the third was propranolol exposure and included data immediately after and 10 min after addition of propranolol. The appropriately trained detector was then used to extract DEs from the data.

In order to estimate s and Σ, 20 high-SNR DEs ($N = 800 \simeq 0.48$ s), as well as 20 noise-only data samples, were identified manually for each detection condition (control, isoproterenol, and propranolol). Analysts identified between 81 and 142 DEs in the control data, depending on the analyst (Table 2.7). Therefore, the 20 control training DEs represent 14-25% of the total control DEs. Similarly, the 20 isoproterenol training DEs represent 15-16% of the total isoproterenol DEs, and the 20 propranolol training DEs represent 24% of the total propranolol DEs. These training DEs were then aligned to their peak values, and averaged to construct s for each detection condition. The identified training noise samples were subdivided into windows ($N = 800 \simeq 0.48$ s), and auto-covariance sequences, $r(k)$, were then calculated at lags $k \in \{-399, -398, \cdots, 399\}$ for each noise window. The total available noise in each detection condition is difficult to quantify, however we can present the training noise for each detection condition in terms of a percentage of the full data for that detection condition. The training noise samples in the control condition, isoproterenol condition, and propranolol condition represent 14%, 7%, and 15% of the control, isoproterenol, and propranolol data, respectively. The Σ for each detection condition was generated by averaging the auto-covariance sequences across that detection condition's noise windows. Both a full noise covariance model (Sec. 2.2.1.2) and a white Gaussian noise covariance model (Sec. 2.2.2.2) were tested.

The size of s and each noise window was empirically selected as $N = 800 \simeq 0.48$ s to ensure that most DEs were captured in full, although some data sets did exhibit both wider and slimmer DEs depending on the drug treatment. The number of noise windows for a given

Table 2.2: Training sample information for each drug treatment. Average training DE SNR is the average SNR of the 20 DEs selected for training for a given drug treatment. The number of windows the training noise sample for each data set could be split into is also listed.

Training Condition	Average Training DE SNR	Number of Training Noise Windows
Control	3.77	34
Isoproterenol	148.07	17
Propranolol	158.97	37

drug treatment varied depending on the length of available noise-only segments in the pre-processed data. Likewise, the shape and SNR of DEs used for estimating s varied between drug treatments. Examples of s from both propranolol and isoproterenol drug treatments, as well as the control, are shown in Fig. 2.3. The average SNR of the training DEs, as well as the number of training noise windows available for each drug treatment is listed in Table 2.6.

2.3.2.3 Results

A total of 7 data traces, from 5 cell cultures, were tested. The first 5 conditions are a control (cell culture 1), immediately after addition of isoproterenol (cell culture 2), 10 min after addition of isoproterenol (cell culture 2), immediately after addition of propranolol (cell culture 3), and 10 min after addition of propranolol (cell culture 3). These conditions will now be referred to as Control-1, Iso-1 0min, Iso-1 10min, Pro-1 0min, and Pro-1 10min, respectively. In order to better test the MaD detector in low-SNR environments we also provide results for a low-SNR control data case (cell culture 4), as well as a low-SNR 15 min after addition of isoproterenol case (cell culture 5). These two conditions will now be referred to as Control-2 and Iso-2 15min, respectively.

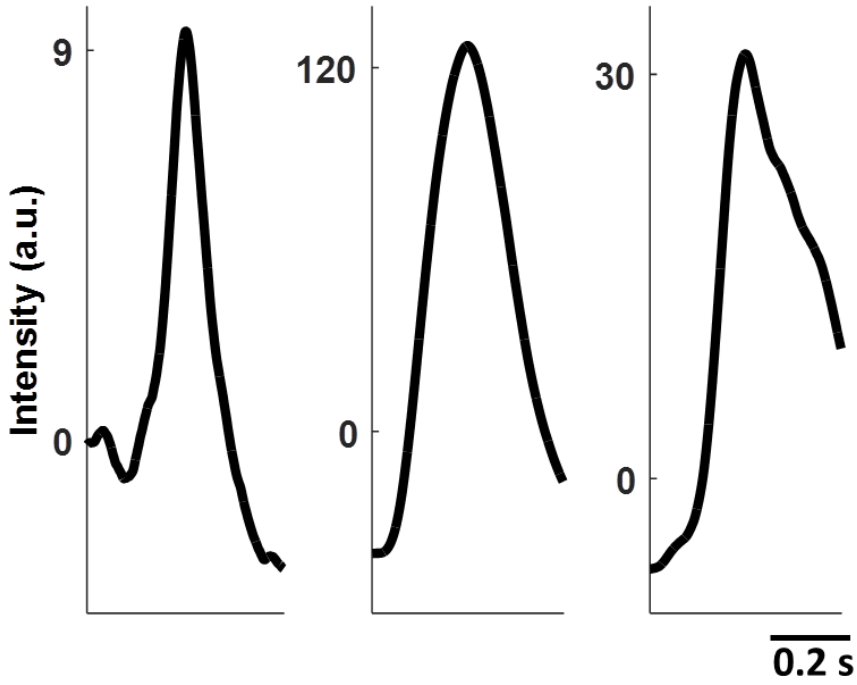

Figure 2.3: Examples of training templates s from control (*left*), isoproterenol (*middle*), and propranolol (*right*) data. Each template is 0.48 s long.

Manual Depolarization Event Identification Three trained analysts identified DE peak times in all 7 presented data traces. The number of identified DEs as well as the average DE SNR for each analyst and each data trace is shown in Table 2.7. The SNR of a given DE was calculated as $\text{SNR}_{DE} = E(DE^2)/\sigma^2_{noise}$, where E is the expectation operator, and σ^2_{noise} was calculated from the 20 noise samples manually selected from each given data trace during training. Each analyst's percentage of agreement with unanimously identified DEs is also presented as a global measure of analyst consistency.

The average SNR of the DEs identified in the Control-1 trace was 5.67. The unanimously identified Control-1 DEs had a slightly higher SNR of 5.71. The average analyst agreement with unanimous Control-1 DEs was 98.87% (only 1 DE was not unanimous). The average SNRs for the Iso-1 and Pro-1 data sets were an order of magnitude higher than the Control-1 DEs. Iso-1 DEs identified by analysts had an average SNR of 69.74, and unanimous Iso-1 DEs had a slightly higher SNR of 70.58. Here the average analyst agreement was 99.03%

(only 2 DEs were not unanimous). Analysts were unanimous about 100% of the Pro-1 DEs and the average Pro-1 DE SNR was 30.97. Examples of the DEs identified by the analysts in each of these drug treatments are shown in Fig. 2.4

The average SNR of the DEs identified in the Control-2 trace was 3.07. The unanimously identified Control-2 DEs had a slightly higher SNR of 3.19. Given this relatively low SNR, the average analyst agreement with unanimous Control-2 DEs was only 75.52%. The average SNR of the Iso-2 15min trace was even lower, at 0.63. Unanimously identified Iso-2 15min DEs had an SNR of 0.65. Similarly, the average analyst agreement with unanimous Iso-2 15min DEs was very low, at only 68.83%. Examples of the DEs identified by the analysts in both of these low-SNR conditions are shown in Fig. 2.5

MaD Detector Performance A total of 5 detection cases were considered: Control-1, Iso-1, Pro-1, Control-2, and Iso-2, where the last 2 (Control-2 and Iso-2) represent specially selected low-SNR data. Unanimous DEs between all 3 analysts were used as the ground-truth for assessing MaD performance. If a detected DE was within 0.12 s, or one quarter of the length of a typical DE, of the true peak time, it was considered a true positive (TP). Otherwise, the detected DE was considered a false positive (FP). Examples of detected DEs are shown in Figs. 2.4 and 2.5.

A true positive vector tp was used to represent all of the ground-truth DEs in a given detection case. If a given DE was successfully detected (TP) it was assigned a value of 1, and 0 otherwise. Similarly, a false positive vector fp was used to represent all of the detected DEs in a given detection case, with each detected DE assigned a value of 1 if the DE was a FP, and 0 otherwise. The means of tp and fp, representing TP and FP rates respectively, were then calculated at 45 incrementally increasing thresholds for each detection case. The TP and FP rates at each threshold were then used to generate ROC curves for each detection case (Figs. 2.4 and 2.5).

Table 2.3: Number and average SNRs of manually identified DEs for each analyst and each data trace. The last column shows the number and average SNR of DEs unanimous among all 3 analysts. The percentage agreement between each analyst and the unanimous DEs is also shown.

Treatment		Analyst 1	Analyst 2	Analyst 3	Average	Unanimous DEs
Control-1	# DEs	60	59	60	59.67	59
	Avg. SNR	5.64	5.71	5.64	5.67	5.71
	Agreement (%)	98.31	100	98.31	98.87	
Control-2	# DEs	81	142	94	105.67	76
	Avg. SNR	3.12	2.91	3.18	3.07	3.19
	Agreement (%)	92.18	53.52	80.85	75.52	
Iso-1	0 min (# DEs)	63	63	62	62.75	62
	10 min (# DEs)	66	67	67	66.50	66
	Avg. SNR	69.76	69.32	70.15	69.74	70.58
	Agreement (%)	99.22	98.46	99.22	99.03	
Iso-2	15 min (# DEs)	43	107	46	65.33	38
	Avg. SNR	0.70	0.58	0.61	0.63	0.65
	Agreement (%)	88.37	35.51	82.61	68.83	
Pro-1	0 min (# DEs)	55	55	55	55	55
	10 min (# DEs)	29	29	29	29	29
	Avg. SNR	30.97	30.97	30.97	30.97	30.97
	Agreement (%)	100.00	100.00	100.00	100.00	

The threshold at which detector performance is closest to TP = 100% and FP = 0% is the optimal threshold and constitutes the best detector performance for that detection case. All following performance metrics are presented as TP or FP Rate [95% confidence interval]. The best performance for the Control-1 case was TP = 97.92 [92.15, 100.00]% and FP = 1.75 [0.00, 5.26]% at a threshold of 4 standard deviations above the noise mean. The best performance for the Iso-1 case was TP = 100.00 [100.00, 100.00]% and FP = 0.78 [0.00, 2.33]% at a threshold of 8 standard deviations above the noise mean, and the best performance for the Pro-1 case was TP = 98.81 [96.43, 100.00]% and FP = 1.19 [0.00, 3.57]% at a threshold

of 6 standard deviations above the noise mean. Overall, the MaD detector performed on par with human analysts and accurately identified DEs at SNR levels of 5 and above.

Performance decreased in the low-SNR detection cases. The best performance for the Control-2 case was TP = 72.37 [61.84, 81.58]% and FP = 16.67 [7.58, 25.76]% at a threshold of 4 standard deviations above the noise mean. The best performance for the Iso-2 case was TP = 65.79 [49.99, 81.59]% and FP = 13.79 [0.44, 27.14]% at a threshold of 2 standard deviations above the noise mean. Note that the training used for detection in the Control-2 and Iso-2 cases was derived from the Control-1 and Iso-1 data sets, respectively. The low analyst agreement, ranging from 76% to 69%, in these low-SNR detection cases made it very difficult to reliably select training samples from the Control-2 and Iso-2 data traces. Therefore, training from the higher SNR data traces had to be used instead. If training data could be reliably selected from the Control-2 and Iso-2 data traces, it is likely that the lower quality of the training samples would adversely affect detection performance.

White Gaussian Noise vs Fully Colored Noise The full noise covariance extracted from the training data (Sec. 2.2.1.2) is a very accurate estimate of the noise parameters. However, if the training noise sample is too small the full noise covariance could be poorly estimated and it may be advantageous to make the simplifying assumption that the noise is white (Sec. 2.2.2.2). We therefore tested the effects of employing the WGN assumption on the data presented in this work. All subsequent results are presented at best performance as (TP% / FP%).

Although the WGN assumption is a less accurate representation of the noise parameters than the full noise covariance, performance under the WGN assumption was not significantly affected in the Iso-1 (100.00%/0.78%) or Pro-1 (97.62%/2.38%) detection cases (ROC curves not shown). The Control-1 case, suffered slightly in performance (93.22%/5.17%), but is still within the previously quoted confidence interval. The Control-2 case showed a more marked

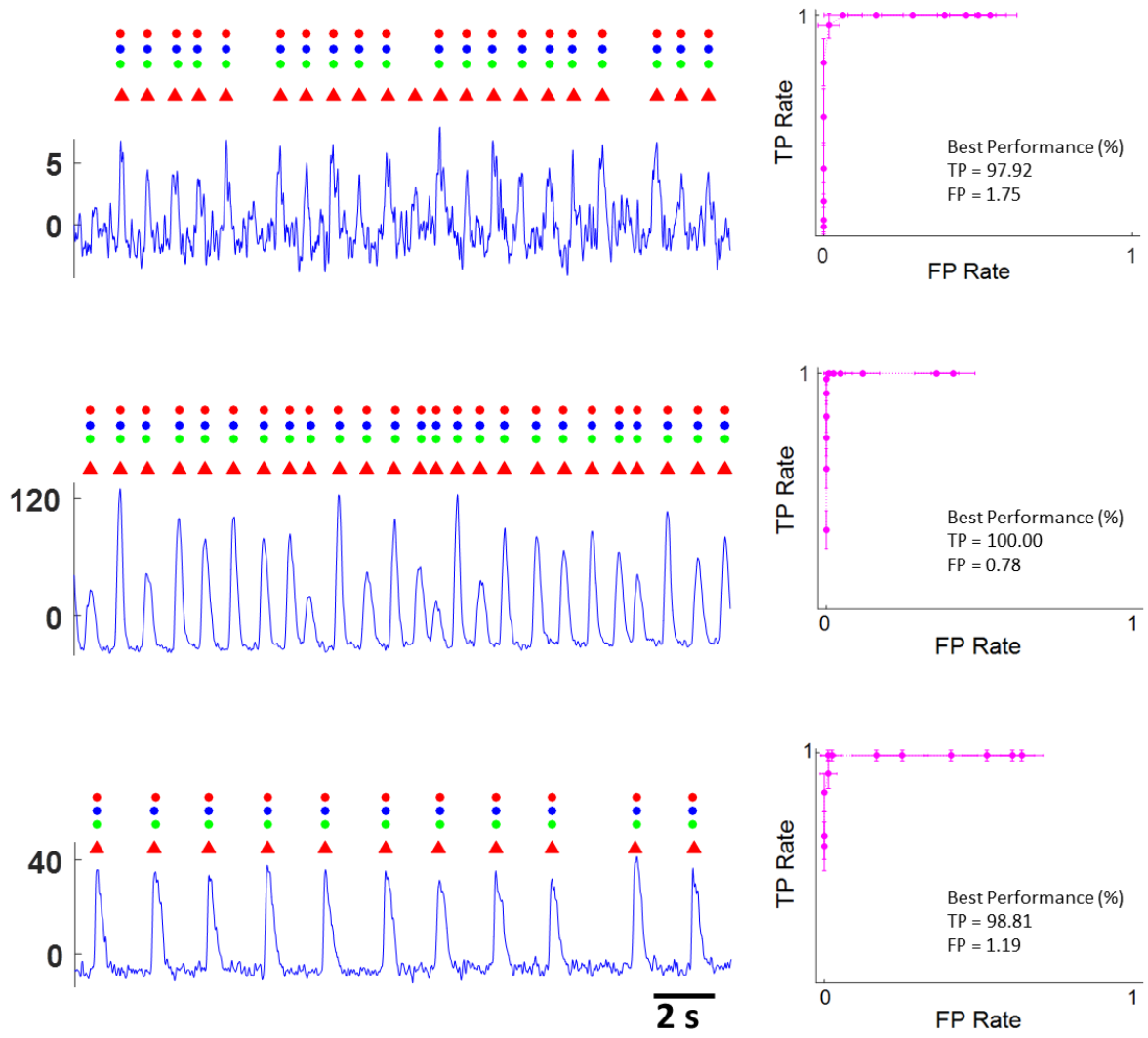

Figure 2.4: Detector Performance. The figure consists of three pre-processed intensity traces from the control (*Top*), isoproterenol (*Middle*), and propranolol (*Bottom*) drug treatments. The circles above each data trace represent the DEs identified by analysts 1-3 (bottom to top respectively). The triangles above each data trace represent MaD detected DEs, using the optimal thresholds identified in the corresponding ROC curves to the right. The optimal threshold is identified as the one resulting in the detector performance closest to TP = 100% and FP = 0%. The error bars represent 95% confidence intervals. The optimal performance for each drug treatment case is presented under each ROC curve. The MaD detector performed with a TP rate of 98-100% and a FP rate of 1-2% for all 3 drug treatments.

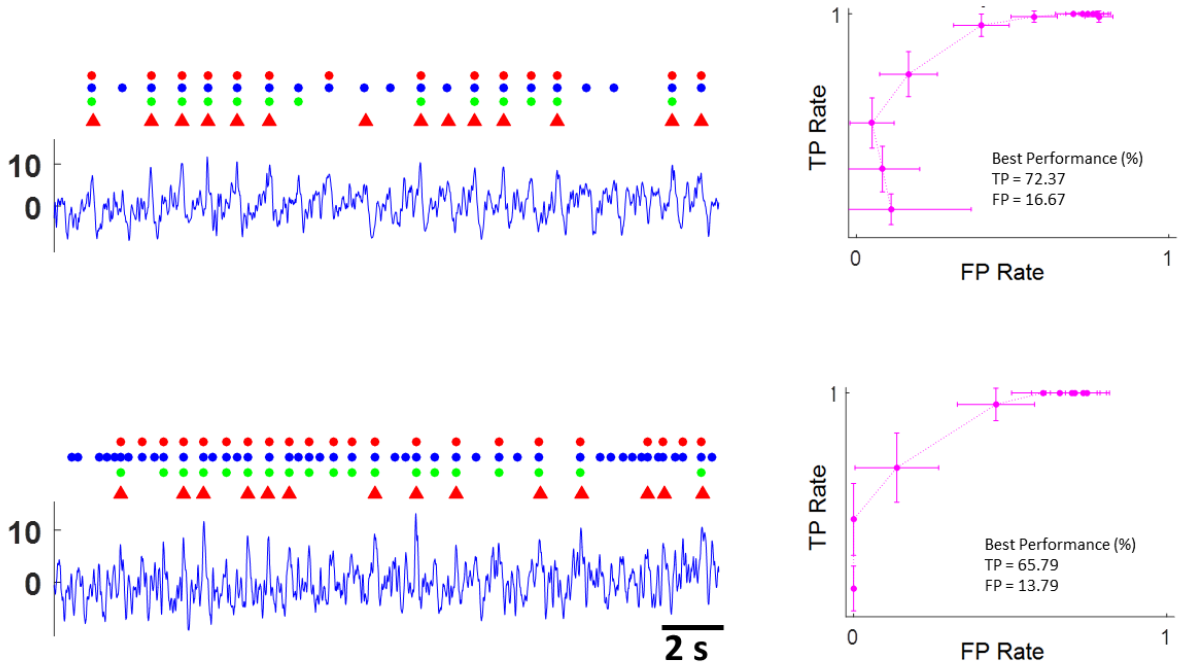

Figure 2.5: Detector Performance At Low SNR. The figure consists of two pre-processed intensity traces from the low SNR control (*Top*), and the low SNR isoproterenol (*Bottom*) drug treatments. The circles above each data trace represent the DEs identified by analysts 1-3 (bottom to top respectively). The triangles above each data trace represent MaD detected DEs, using the optimal thresholds identified in the corresponding ROC curves to the right. The error bars represent 95% confidence intervals. The optimal performance for each drug treatment case is presented under each ROC curve. As the SNR decreased, the performance of the MaD detector also decreased. Performance was still adequate for SNR $\simeq 3$, but not for SNR ≤ 1.

decrease in detection efficacy (85.53%/32.29%). Curiously, the Iso-2 case, which has the lowest SNR of all of the data presented in this work, showed an increase in both the best TP rate, and the best FP rate (89.47%/34.62%), which resulted in a comparable overall performance.

These results imply that the matched filter under a WGN assumption is sufficient for DE detection at SNR levels of 5 and higher, but may lead to a decrease in detection efficacy at lower SNRs.

Effects of Differing Training Samples For some applications, it may be advantageous to reduce the training samples size, in the interest of time. In our experience, the noise sample is highly dependent on the length of individual noise segments identified by the user. If long noise segments (more than 10 times the window size) are available, then as little as 10 segments can be sufficient for accurate DE detection. Similarly, the signal sample is dependent on SNR. If high SNR (SNR $\geq$ 5) DEs are available for training, then as little as 10 samples is sufficient for accurate DE detection. However, if the signal SNR is lower, then at least 20 samples should be used to maximize detection efficacy. Note that data pre-processing has little effect on SNR and serves only to remove large scale artifacts produced by the motion of the beating cell membrane and photobleaching of the VSD. Therefore variations in pre-processing methods should not affect the training paradigm.

It may also be useful to be able to re-use training samples between different data sets. To determine how re-using training samples from one data set to detect DEs from a different data set affects detection performance, we compared detection performance when a native training sample (derived from the data being tested) is used with detection performance when a non-native training sample (derived from data of a different cell membrane than the one being tested) is used. For this analysis we only tested the Control-1, Iso-1, and Pro-1 detection cases, as the low-SNR cases may not lead to reliable enough conclusions. All subsequent results are presented at best performance as (TP% / FP%). Using a non-native isoproterenol training sample on the Iso-1 data resulted in (96.88%/4.62%) using the full noise covariance, and (99.22%/0.78%) under the WGN assumption (ROC curves not shown). Similarly, using a non-native propranolol training sample on the Pro-1 data resulted in (90.48%/9.52%) using the full noise covariance, and (100.00%/0.00%) under the WGN assumption. In the Control-1 case, using a non-native control training sample resulted in (98.31%/1.69%) using the full noise covariance, and (96.61%/1.72%) under the WGN assumption. The detector worked well even when Iso-1 training was used on Pro-1 data (85.71%/26.53% using full noise covariance; 100.00%/0.00% under the WGN assumption)

and vice versa (79.69%/28.67% using full noise covariance; 98.44%/1.56% under the WGN assumption).

In all of these cases we see that detection performance is maintained if a non-native training sample is used. Furthermore, efficacy is most often preserved under the WGN assumption. When the full noise covariance is used with a non-native training sample, detection performance may begin to degrade, indicating that in the absence of a native noise covariance, the filter may perform better using a WGN approximation than using a well estimated but inaccurate, non-native noise covariance.

2.3.2.4 Discussion and Conclusion

VSDs address limitations of existing methods for measuring hiPS-CM DEs. Patch clamping offers precise measurement of transmembrane voltages, but is invasive in nature, requiring destructive membrane puncture to position electrodes that prevents longitudinal experiments. Only single cells may be assessed using a complex apparatus [5, 37, 111]. The growing field of organ-on-a-chip platforms demands non-invasive fluorescence based *in situ* endpoints for assessing hiPS-CM [35, 67]. VSD fluorescence intensity measurements using 2-photon microscopy are noninvasive, non-destructive, and allow for longitudinal electrophysiological assessment of live, 3D, cardiac tissues in microfluidic-based devices. Quantitative assessment of the DEs of actively beating cardiomyocytes is critical to understanding the electrophysiology underlying these cardiac tissue models.

The MaD detector presented in this work was developed to detect DE from VSD imaged cardiomyocytes. Although some detection algorithms exist for similar use in calcium imaging applications, they are either not tailored well to VSD data [97, 8, 3], or have not been tested on cardiomyocyte data [50, 112]. The MaD detector provides a tool that is specifically designed for automatic detection of VSD imaged cardiomyocyte DEs. It has proven to work

exceptionally well, especially on data with SNR ≥ 5, with TP rates ranging from 98-100% and FP rates ranging from 1-2%. This level of performance is largely maintained even if a simplified noise covariance model (WGN) is used, and if a non-native training sample is used. Furthermore, unlike other VSD-approaches that use pre-defined waveform features, the MaD detector uses a data-driven sample of the entire waveform to detect DEs, resulting in a non-biased DE selection criterion that can accurately detect waveforms. Furthermore, the MaD detector's data driven design means it can be easily modified to assess the electrophysiology of other excitable cell populations [98] and data types [104].

Although the DEs presented here appear periodically, the MaD algorithm takes a generalized approach to signal detection. This allows for the detection of irregular DEs that may be indicative of drug-induced side effects. Early after depolarization (EAD) and delayed after depolarization (DAD) events are potentially lethal and may occur at irregular intervals [23, 113]. MaD does not incorporate the expected periodicity of cardiac DEs into the detection algorithm, allowing for the flexibility to detect these types of events. In the event that periodic signals are explicitly desired at low-SNR, the MaD detector could be modified to apply a variable threshold to pick up low-SNR DEs at the expected time intervals, without increasing the number of FPs throughout the entire dataset. Future work with MaD will focus on variability from DE to DE to detect these rare, but potentially lethal EADs and DADs.

Overall the MaD detector is a useful new tool for the study of cardiomyocyte electrophysiology. Combined with the use of voltage-sensitive dyes, it allows for non-invasive, image-based, and automated detection of cardiac DEs. This study demonstrates the ability of this tool to detect DEs across a wide range of SNRs and various levels of model complexity.

Acknowledgements I would like to acknowledge Amin Gosla for his assistance in manually detecting DEs.

2.3.3 Matched filter for Multi-unit Calcium Event detection (MMiCE)

The work presented here is adapted from Szymanska et al. 2016 [104]. Briefly, we developed a Matched filter for Multi-unit Calcium Event (MMiCE) detection to extract calcium events (CEs) from fluorescence intensity traces of simulated and experimentally recorded neuronal calcium imaging data. MMiCE performed with a true positive (TP) rate of 98.27% (with a 95% confidence interval of [96.89, 99.65]%), and a false positive (FP) rate of 6.59% (with a 95% confidence interval of [4.03, 9.15]%) on simulated data with SNR 0.2. It reached perfect performance on simulated data with SNR $\geq$ 2. MMiCE performance on real data was also very promising (TP = 100 [100.00, 100.00]%, FP = 2.04 [0.00, 6.14]%). Trained analysts indicated the algorithm was more consistent and reliable than manual CE tagging. Detection time with MMiCE was reduced from roughly 24 hours per data set manually, to just 75-s per data set on average (0.0047-s per region of interest per second of recording). Overall, the MMiCE detector greatly reduced run time compared with manual CE detection, and performed exceptionally well on both simulated data at SNR as low as 0.2, and experimentally recorded neuronal calcium imaging data. The MMiCE detector is accurate, reliable, and well suited for wide-spread use.

2.3.3.1 Background on Calcium Event Detection

Functional multineuron calcium imaging (fMCI) has been particularly successful as a means of studying the interactions of multiple (often hundreds) of neurons in large populations [106] with specific applications both *in vitro* [12, 40, 86, 87, 88], and *in vivo* [47, 105]. Regardless of the details in fluorescence imaging, or focus on either synaptic or somatic calcium waves, calcium imaging in neurons often suffers from low SNR. This makes signal detection very difficult, prompting the development of many automated or partially automated detection approaches.

Pioneers in the field first represented calcium imaging data as fluorescence changes ($\Delta F/F$), and established a link between somatic calcium transient waves and action potentials (APs) [92, 47]. Subsequent studies linked the amplitude of calcium transient waves to AP firing patterns and rates [119, 66, 27]. Although it is now universally accepted that relative amplitudes of fluorescent CEs are indicative of the frequency or number of APs fired by the cell, algorithms back-calculating AP firing rates are difficult to generalize. The amplitude of a given CE depends on the amount of fluorescent calcium indicator taken up by the cell, as well as the proximity of the region of interest (ROI) to the focal imaging plane. When performing imaging population studies, such as those using fMCI [106], it is impractical and often impossible to calculate the exact distance from the focal plane and to quantify the amount of indicator present in each cell or ROI. Comparison of CE amplitudes across ROIs or data sets is therefore unreliable. Furthermore, low-SNR signals where CEs are generated from 5 or fewer APs often fail to be detected [66]. When they are detected, the low-SNR CEs often lack the necessary resolution for AP firing rate estimation. Although Sasaki et al. overcame many of these issues with their support vector machine approach, they still fell victim to rapid performance degradation at low SNR [88]. It may therefore be more advantageous to sacrifice precision in AP firing rate estimation for better accuracy in CE detection at low SNR.

Other approaches to CE detection less focused on AP firing rates include thresholding the time derivative of fluorescent changes in given ROIs [81, 40], Hanning filters [12], and matched filters [47]. These approaches work well in high-SNR settings (SNR ≥ 10) but their performance is either not reported at low SNRs or quickly degrades. Since a large fraction of CEs fall below SNR 2, there is a pressing need to develop automated methods for low-SNR CE detection. Some denoising algorithms have been developed to enhance the SNR of calcium imaging data [43, 60]. However, they are not applied or tested as detection tools, but rather as post hoc analyses. Due to the various shortcomings of these methods, as well as availabil-

ity and implementation hurdles, manual CE detection remains the primary means of scoring calcium imaging data even though it is both prone to bias and incredibly laborious.

Our work focuses on the development of a Matched filter for Multi-unit Calcium Event (MMiCE) detection specifically tailored for low-SNR CEs. Unlike the matched filter presented by Kerr et al. [47], the algorithm developed here is completely data driven. It is also applied and tested in environments with SNR as low as 0.2, well below the range of existing algorithms. Although only tested on fMCI data here, due to its data-driven design, the algorithm is easily generalizable to other fluorescent neural imaging modalities, such as voltage-sensitive-dye imaging [103].

2.3.3.2 Methods

Data Collection MMiCE was tested under 3 paradigms: (1) simulated somatic calcium imaging data (ground-truth available), (2) experimentally recorded simultaneous somatic fMCI and patch-clamp data (ground-truth available), as well as (3) experimentally recorded, from now on referred to as real, somatic fMCI and dendritic spine fMCI data (no ground-truth available).

The full data collection procedure for paradigms (2) and (3) is in Appendix A.2 as well as Szymanska et al. 2016 [104]. Briefly, *ex vivo* organotypic slice cultures were prepared from the hippocampus of 7 day old Wistar/ST rats [105]. These cultures were used for simultaneous somatic fMCI and patch-clam recordings (2), as well as dendritic spine fMCI recordings (3). For the simultaneous somatic fMCI and patch-clam (fMCI-PC) recordings the cultures were incubated in an OGB1 dye solution, washed, and mounted in a recording chamber. CA3 pyramidal neurons selected for recording were then voltage-clamped at 0 mV. For the dendritic spine fMCI recordings, CA3 pyramidal neurons selected for recording were voltage-clamped at -30 mV and loaded with a Fluo-4 solution. Lastly, 400 μm horizontal

hippocampal slice of 3 week old C57Bl/6J mice [108, 74] were used for somatic fMCI recording. The slices were locally loaded into the CA1 stratum pyramidale using a pipette filled with an OGB1 dye solution. Fluorophores were excited at 488 nm with a laser diode and visualized using a 507 nm long-pass emission filter. Videos were recorded at 10 Hz for the fMCI-PC data and 50 Hz for all other fMCI data.

In total, one data trace was acquired for fMCI-PC recordings. The recording measured spontaneous activity from a CA3 pyramidal neuron from the *ex vivo* slice cultures. Three somatic fMCI data sets were acquired measuring spontaneous activity from CA1 pyramidal neurons from the acute slices. Lastly, five dendritic spine fMCI data sets were acquired measuring spontaneous synaptic inputs from CA3 pyramidal neurons from the *ex vivo* slice cultures.

ROIs for all somatic and dendritic fMCI data sets were identified manually using a custom software [40]. The average fluorescence for each ROI was then calculated and used to determine the change in fluorescence $\Delta F/F = (F_1 - F_0)/F_0$, where F_1 is the fluorescence intensity at any time point, and F_0 is the average baseline fluorescence intensity 1-s before and after F_1. This normalization with a 2-s epoch around the target frame is needed to compensate for photobleaching. The resulting fluorescence intensity trace, X, for each ROI was then used for further analysis.

Somatic and dendritic spine fluorescence intensity traces, X, were independently plotted and CE peak times were manually identified by six trained human analysts. CEs were not manually identified for fMCI-PC data.

Generating Simulated Data In order to assess the MMiCE detector's performance at varied SNR levels and to establish a ground truth for comparison, we simulated 8 fluorescence traces, each with a different average SNR. To capture as much of the real data character-

istics as possible, the simulated traces, $X_{\text{Sim},i}$ where $i \in \{1, 2, \cdots, 8\}$, were generated using fluorescence traces from the calcium imaged neuronal somas described in Sec. A.2.4.

Four different CE shapes were identified manually by human analysts from three somatic fMCI data sets (Sec. A.2.4). 20 examples of each shape were then selected from among the three data sets, averaged, low-pass filtered at 5 Hz, and normalized to construct 4 distinct CE templates (Fig. 2.6(B)). The low-pass filter was necessary because averaging did not sufficiently smooth the CE templates. The CE templates, s_c, $c \in \{1, 2, 3, 4\}$, were cropped to lengths of 1-s, 1.3-s, 1.2-s, and 1.3-s respectively.

Similarly, 80 samples of varying length noise-only data were manually selected from the three somatic fMCI data sets. Each sample was normalized to be zero mean and unit variance. The noise samples were then all concatenated into a single time series of length $T_{noise} = 89000 \simeq 29.7$min at a sampling rate of 50 Hz. The four CE templates were superimposed with the noise at random time points such that no spikes overlapped, and there were 100 CEs of each spike shape for a total of 400 CEs.

The SNR level was controlled by scaling the sizes of the CE templates being added to the noise. Here we define SNR as

$$SNR = \frac{E((f \cdot s)^2)}{\sigma^2_{noise}} = f^2 E(s^2) \tag{2.9}$$

where E is the expectation operator, $s \in \mathbb{R}^{1\times N}$ is the CE template, f is a scaling factor, and $\sigma^2_{noise} = 1$ is the noise variance. We can then solve for the scaling factor $f_{i,c}$ for each SNR_i and each CE template s_c.

The fluorescence traces were simulated at SNR levels of 0.2, 0.5, 1, 2, 4, 6, 9, and 14, where SNR 1 indicates that the signal has the same power as the noise. The scaling factor for each template at each SNR level is listed in Table 2.4. Figure 2.7(A) shows a 30-s excerpt

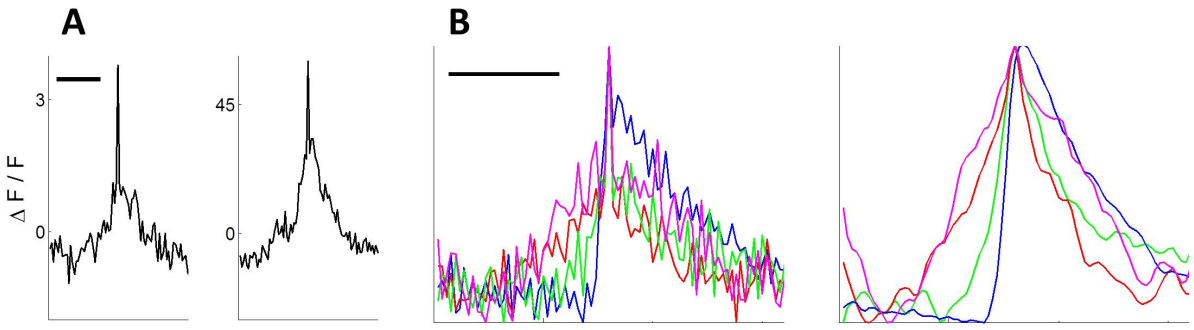

Figure 2.6: (A) Examples of training templates s from (*Left*) somatic fMCI data set 1 and (*Right*) dendritic spine fMCI data set 1. The scale bar corresponds to 0.5-s for both figures. Note that the dendritic spine template is wider and has a much higher amplitude than the somatic template. (B) CE templates used for generating simulated data, with s_1, s_2, s_3, and s_4 in blue, red, green, and magenta respectively. (*Left*) Normalized and unfiltered templates. (*Right*) Normalized and filtered (5Hz low-pass) templates. The scale bar corresponds to 0.5-s for both figures.

containing 6 simulated CEs from each simulated data trace. Note that the only difference between each simulated data trace is the SNR of the CEs; the noise and CE times remain constant.

Filter Training The MMiCE detector is completely data driven, allowing it to maintain high performance even if large variations exist between data sets. Both s and Σ from Eqs. 2.1 - 2.3 are estimated from the data. First, 20 high-SNR CEs (N = 1.2-s for real data, and N = 1.6-s for simulated data) were identified manually, aligned to their peak values, and averaged to estimate s for each data set. In order to ensure that the entire data set is well represented, the CEs were always selected from multiple different ROIs (when possible), all of which exhibit high SNR. The CEs were also selected to match the traditional sharp rise and slow decaying shape of a calcium transient wave as much as possible. The specific SNR values and CE shapes available for estimating s varied between data sets. Examples of s from both somatic fMCI and dendritic spine fMCI data are shown in Fig. 2.6(A). Due to the

Table 2.4: Scaling factor f for each template at each SNR level

		Scaling Factor f			
		s_1	s_2	s_3	s_4
SNR Level	0.2	0.83	1.13	1.01	0.87
	0.5	1.30	1.79	1.60	1.37
	1	1.85	2.53	2.26	1.94
	2	2.61	3.58	3.20	2.74
	4	3.69	5.07	4.53	3.88
	6	4.52	6.20	5.55	4.75
	9	5.53	7.60	6.79	5.82
	14	6.90	9.48	8.47	7.26

limited CEs available in the single fMCI-PC data trace acquired, only 10 CEs were selected for training this data.

Next, 20 noise-only samples of various lengths, from multiple different ROIs, were identified manually for each data set. The noise samples were then subdivided into noise windows ($N = 1.2$-s for real data, and $N = 1.6$-s for simulated data). The number of noise windows varied between real data sets depending on the lengths of the 20 noise samples selected for training. The training noise samples were identical among all simulated data sets. Any data points remaining after the subdivision were discarded. Auto-covariance sequences, $r(k)$, were then calculated at lags $k_{real} \in \{-0.58\text{-s}, -0.56\text{-s}, \cdots, 0.58\text{-s}\}$ for real data sets and $k_{sim} \in \{-0.78\text{-s}, -0.76\text{-s}, \cdots, 0.78\text{-s}\}$ for simulated data sets, for each noise window. The sequences were averaged across all noise windows for a given data set, and used to generate Σ for that data set.

The size of s and each noise window was empirically selected as $N = 1.2$-s for real data, and $N = 1.6$-s for simulated data. The real data window, $N = 1.2$-s, ensured that most somatic and spine CEs were captured in full, although some data sets did exhibit both wider and slimmer CEs (independently of data type). The somatic CEs used to generate the simulated data set were specifically selected to have a high SNR. These CEs tended to

be wider than the average CE, and therefore the window size for the simulated data was increased to $N = 1.6$-s. These window sizes are consistent with known typical CE durations for AP bursts in neurons [47, 92].

Setting Detection Thresholds The 8 fMCI data sets used in this study (excluding the fMCI-PC data trace) contained 179 ROIs on average, ranging from 75 to 313 ROIs. Because calcium indicator concentration is not uniform across all ROIs, and the independent ROIs may not all perfectly be in the focal plane of the camera, the resulting fluorescence intensity traces exhibit varying amplitudes as well as varying SNRs. Therefore a single filter threshold cannot be set for the entire data set. Similarly, it would be too laborious to manually set the filter threshold for each ROI in a data set. It is therefore necessary to devise an automated thresholding algorithm that can account for varying SNRs and amplitudes throughout a given data set.

To accomplish this goal, threshold values were calculated independently for each fluorescence intensity trace X_r, where $r \in \{1, 2, \cdots, R\}$, and R is the total number of ROIs in the data set. The thresholding parameters are calculated from the filtered signal which contains both CEs and noise (Eqs. 2.4- 2.5). This thresholding approach generates thresholds that can accommodate varying signal amplitudes and SNRs for each X_r, and allows the user to set a specific sensitivity value, a, for the entire data set.

Performance Analysis CEs detected for each of the simulated data traces, as well as those detected from the fMCI-PC data trace, were compared against the ground-truth spike times. In the case of the fMCI-PC data, the ground-truth spike times were identified as the peak times of the first APs in an AP burst, where a single AP burst consisted of APs firing within 0.1-s of each other. The first 3.3 min of the simulated data traces were used for training, and the remaining test data (26.4 min) contained 346 CEs. Similarly, 121-s of the

fMCI-PC data trace was used for training and the remaining test data (179-s) contained 48 AP bursts. A detected CE was considered a true positive (TP) if the detected peak time was within 0.8-s of the true peak time. This corresponds to half of the length of the simulated CEs. Otherwise, the detected CE was considered a false positive (FP). Vector tp was used to represent all of the ground-truth CEs, with each index assigned a value of 1 if the CE was detected (TP) and 0 otherwise. Similarly, vector fp was used to represent all of the detected CEs, with each index assigned a value of 1 if the CE was a FP and 0 otherwise. TP and FP rates for 100 incrementally increasing thresholds were then calculated as the means of tp and fp, respectively, at each of the thresholds. The false negative (FN) or omission rate can be calculated as (1 − TP Rate). The TP and FP rates at each threshold were then used to generate ROC curves for each simulated data trace.

2.3.3.3 Results

Detection on Simulated Data

Training for Simulated Data The simulated data traces were all treated as independent data sets, and the MMiCE detector was separately conditioned on each set. In general, a real data set will have a wide range of SNR levels, usually with some high-SNR spikes to pick for training. The presence of low-SNR spikes in the training sample may result in a poorly determined template s (Sec. 2.2.1), which would adversely affect detection. However, in this case all of the CEs in each simulated data set have roughly the same SNR. The training for each SNR level was kept separate in order to prove that the MMiCE detector was accurate even if the CE training samples were of poor quality and very low SNR.

The first 3.3 min of the simulated data traces were used for training. Training CEs were selected with prior knowledge of simulated CE peak times. Noise samples for training were

selected manually, and were the same for all $X_{\mathrm{Sim},i}$. The training noise was subdivided into 56 noise windows (N = 1.6-s), and auto-covariance sequences from each window were then averaged to generate the noise covariance matrix Σ_{Sim} (Sec. 2.3.3.2). The data used for training was omitted when assessing MMiCE detection performance.

MMiCE Detector Performance on Simulated Data The simulated data sets were filtered using the MMiCE detector and CE times were identified as the peaks of the filter output above a given threshold. Performance was assessed at 100 incrementally increasing thresholds with sensitivity values, a, ranging from 0.5 - 50.0 (Eq. 2.5). The resulting TP and FP rates were used to generate ROC curves for all 8 simulated data sets (Figs. 2.7(B), 2.8). The MMiCE detector performed exceptionally well on the simulated data, even at SNR levels as low as 0.2. Performance metrics are listed as TP or FP Rate [95% confidence interval].

The MMiCE detector was first applied in its simplest form, assuming WGN statistics (Sec. 2.2.2.2, Eq. 2.3). Under this assumption, the MMiCE detector achieved perfect performance (TP = 100 [100.00, 100.00]%, FP = 0 [0.00, 0.00]%) for SNR $\geq$ 2. At SNR 1 the performance was slightly affected (TP = 100 [100.00, 100.00]%, FP = 0.29 [0.00, 0.86]%), with a further dip in performance at SNR 0.5 (TP = 97.98 [96.84, 99.12]%, FP = 0.88 [0.00, 1.87]%) and SNR 0.2 (TP = 90.46 [87.35, 93.57])%, FP = 12.81 [9.34, 16.28]%). The MMiCE detector's ROC curves, depicting performance at all thresholds under the WGN assumption, are shown for each data set in Fig. 2.7(B). The best performances at each SNR level, and the corresponding threshold sensitivity values are shown in Table 2.5.

To see if even better outcomes could be achieved for SNR 0.2 - 2, the MMiCE detector was then applied in its full complexity, using the full noise covariance to construct the matched filter (Sec. 2.2.1.2). For simulated data sets of SNR 1 and SNR 2, performance was not affected. This is due to the fact that performance was already perfect for SNR 2 and very

Table 2.5: Best MMiCE detector performances, and the associated threshold sensitivity values a for each simulated data set. Both performances under the WGN assumption and using the full noise covariance are shown. Because perfect performance was achieved under the WGN assumption for SNR 4 and above, the full noise covariance detector was not applied to these data sets. This is reflected by the lack of threshold sensitivity values for SNR 4-14 under Full Covariance

		Best MMiCE Detector Performance					
		WGN			Full Covariance		
		TP %	FP%	a	TP%	FP%	a
SNR Level	0.2	90.46	12.81	2	98.27	6.59	3
	0.5	97.98	0.88	3.5	99.71	1.15	4.5
	1	100	0.29	4	100	0.29	7
	2	100	0	4-8	100	0	10-11
	4	100	0	4.5-13	100	0	–
	6	100	0	5-17.5	100	0	–
	9	100	0	8-22.5	100	0	–
	14	100	0	4.5-29	100	0	–

near perfect for SNR 1 under the WGN assumption. Performance was improved for SNR 0.5, with TP = 99.71 [99.14, 100.00]% (from 97.98 [96.84, 99.12]% before), and FP = 1.15 [0.03, 2.27]% (from 0.88 [0.00, 1.87]% before). Although the FP Rate slightly increased, overall this is closer to perfect performance than under the WGN assumption. An even greater improvement was seen for SNR 0.2, with TP = 98.27 [96.89, 99.65]% (from 90.46 [87.35, 93.57])% before), and FP = 6.59 [4.03, 9.15]% (from 12.81 [9.34, 16.28] before). The MMiCE detector's ROC curves, depicting performance at all 100 thresholds when the detector employs the full noise covariance for training, are shown in Fig. 2.8 for simulated data sets of SNR 0.2 - 2. The best performance at each SNR level, and the corresponding threshold sensitivity values for SNR 0.2 - 2 are shown in Table 2.5.

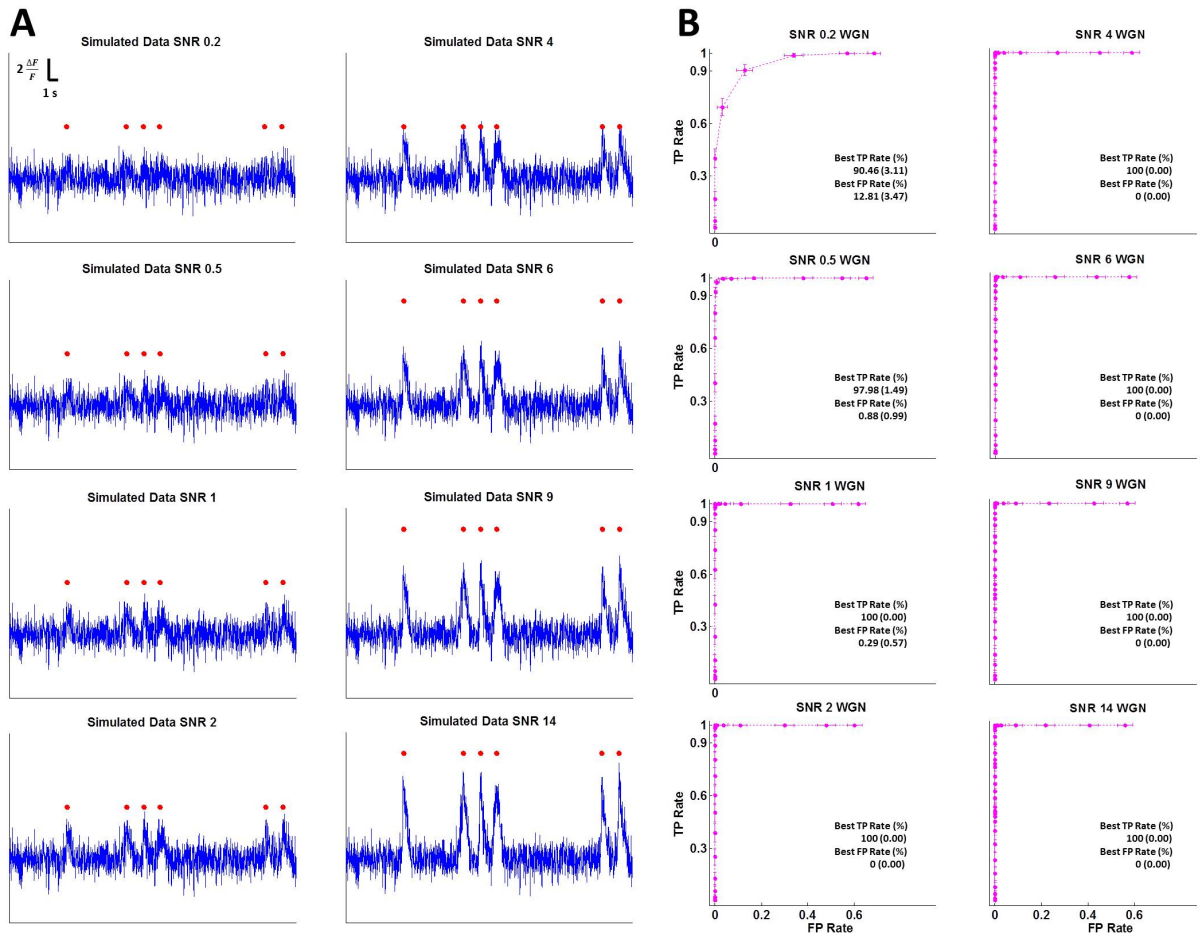

Figure 2.7: (A) 30-s excerpts of simulated data for SNR 0.2, 0.5, 1, 2, 4, 6, 9, and 14. Red circles above each graph represent the peak locations of 6 inserted CE templates. Going from left to right the inserted templates are $s_1, s_4, s_1, s_4, s_1, s_3$. (B) ROC curves for all simulated data sets, across 100 incrementally increasing thresholds, with the MMiCE detector assuming WGN statistics. The error bars represent 95% confidence intervals. The optimal performance for each simulated data set is presented under each curve with 95% confidence errors in parenthesis. Note that perfect performance (TP = 100 [100.00, 100.00]%, FP = 0 [0.00, 0.00]%) is reached for SNR 2 - 14. Performance only begins degrading significantly at SNR 0.2 (TP = 90.46 [87.35, 93.57])%, FP = 12.81 [9.34, 16.28]).

Detection on Simultaneous fMCI and Patch-Clamp Data

Training for fMCI-PC Data A single simultaneous somatic fMCI and patch-clamp data trace was used for this experiment. 121-s of the trace was used for training. Due to the limited amount of data available, only 10 CEs with relatively high SNRs, and smooth

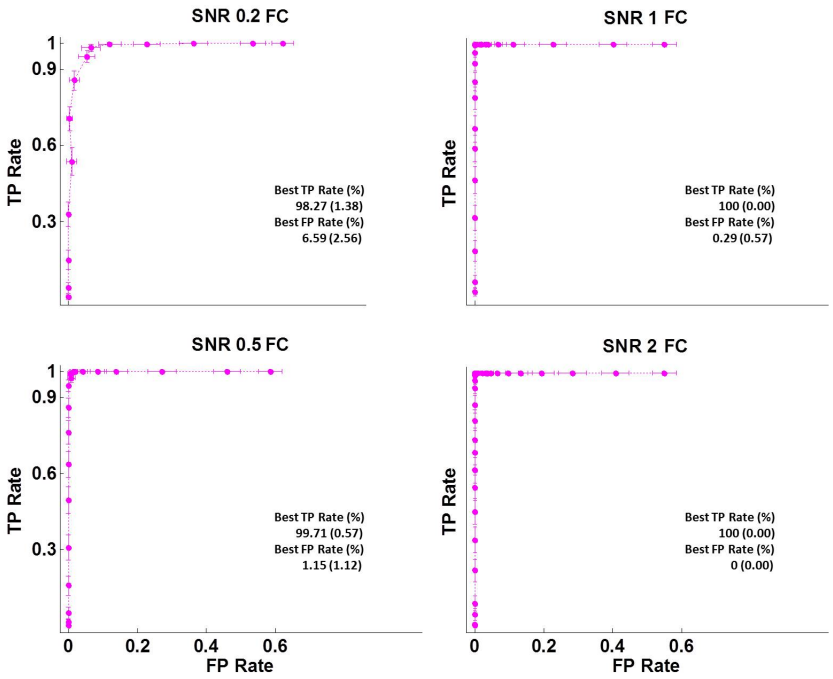

Figure 2.8: ROC curves for simulated data sets of SNR 0.2 - 2, across 100 incrementally increasing thresholds, with the MMiCE detector employing the full noise covariance. The error bars represent 95% confidence intervals. The optimal performance for each simulated data set is presented under each curve with 95% confidence errors in parenthesis. Note that perfect performance (TP = 100 [100.00, 100.00]%, FP = 0 [0.00, 0.00]%) and near perfect performance (TP = 100 [100.00, 100.00]%, FP = 0.29 [0.00, 0.86]%) is still reached for SNR 2, and SNR 1, just as in the WGN case. However, performance improved significantly from the WGN case for SNR 0.5 (TP = 99.71 [99.14, 100.00]%, FP = 1.15 [0.03, 2.27]%) and SNR 0.2 (TP = 98.27 [96.89, 99.65]%, FP = 6.59 [4.03, 9.15]%).

shapes could be manually selected to generate the training template, s. Similarly, the 20 manually selected noise samples were relatively short, therefore only 91 noise windows were available for training (Table 2.6). This under-sampled training set could result in a poorly determined s as well as Σ, and thereby adversely affect detection performance. The average SNR of the training CEs was calculated as

$$SNR_{CE} = \frac{E(CE^2)}{\sigma^2_{Noise}} \tag{2.10}$$

where σ^2_{Noise} was calculated from the training noise samples, and $E(CE^2)$ is the expected value of a CE^2. Note that unlike in Eq. 2.9, the CE in this case contains a noise component.

Therefore the SNRs calculated from the real data will be biased. Due to this bias an SNR of 2 calculated from the real data will roughly correspond with an SNR of 1 calculated for the simulated data. The average SNR of the fMCI-PC training CEs is 8.23 (Table 2.6). All data selected for training were omitted from the performance analysis.

MMiCE Detector Performance on fMCI-PC Data The ground-truth CEs for the fMCI-PC data trace, identified as described in Sec. 2.3.3.2 had an average SNR of 10.54. As explained in Sec. 2.3.3.3, this roughly corresponds with simulated data of SNR 9. As in the simulated data case, the fMCI-PC data trace was filtered using the MMiCE detector under the WGN assumption as well as using the full noise covariance. Performance for both approaches was assessed at 100 incrementally increasing thresholds with sensitivity values, a, ranging from 0.5 to 50.0 (Eq. 2.5). The resulting ROC curves for both the WGN case, and using the full noise covariance are shown in Fig. 2.9 *Left*. Best performance under the WGN assumption was TP = 95.83 [89.97, 100.00]% and FP = 2.13 [0.00, 6.41]%. The corresponding a range was 3.5 to 4. This constitutes a total of 3 errors, 1 false positive, and 2 false negatives or omissions. The false positive was a result of misidentifying an artifact of photobleaching as a CE resulting from cell activity. The first false negative was a failure of the MMiCE detector to identify an AP burst that resulted in fMCI activity of SNR 2.07, which is much lower than average for this data trace, and as explained in Sec. 2.3.3.3 is equivalent to SNR 1 in the simulated data. The second false negative was the MMiCE detector's failure to identify one of two overlapping CEs. The first CE was identified but the second was omitted. Best performance using the full noise covariance was TP = 100.00 [100.00, 100.00]% and FP = 2.04 [0.00, 6.14]% at $a = 2.0$. This constitutes 1 error, which is the same false positive as in the WGN case, i.e. misidentifying an artifact of photobleaching as a CE resulting from cell activity. The two false negatives, however, were resolved when using the full noise covariance. An excerpt of the spike train as determined by patch-clamp recordings, the fMCI data trace, detected CEs, and the filter output using the full noise

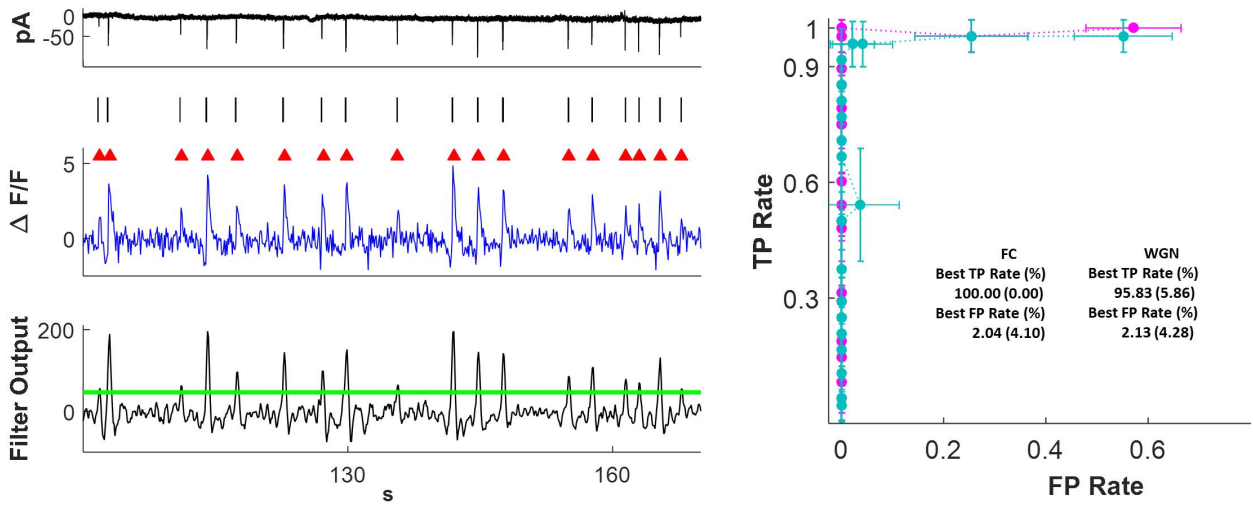

Figure 2.9: MMiCE detector performance and real somatic fMCI data as verified by patch-clamp recordings. (*Left*) The top shows the ground-truth patch-clamp recordings. In the middle is the corresponding ΔF/F data with MMiCE CE peaks detected at the optimal threshold ($a = 3.5$) marked by red triangles. The spike-train derived from the patch-clamp recordings is shown above the detected CEs for easy comparison. The bottom trace shows the filter output using the full noise covariance and the optimal detection threshold as determined by the ROC curves presented on the *Right*. (*Right*) ROC curves for the fMCI-PC data shown on the *Left*, across 100 incrementally increasing thresholds. The two curves depict the MMiCE detector's performance under the WGN assumption (cyan) and when using the full noise covariance (FC, magenta). The error bars represent 95% confidence intervals. Optimal performances with 95% confidence errors in parenthesis are presented under the curves.

covariance are shown in Fig. 2.9. Overall, the MMiCE detector did exceptionally well in CE detection from somatic fMCI data, as verified via patch-clamp recording.

Detection on Somatic fMCI and Dendritic Spine fMCI Data

The CE SNR in fMCI-PC data used for validation of the MMiCE algorithm in Sec. 2.3.3.3 was 10.54. Although CEs of this SNR do appear for some ROIs in many data sets, they are not common. Therefore MMiCE detector performance was also tested in more realistic SNR settings with SNR ranging from 1 to 3 in real fMCI data. Three somatic fMCI and

Table 2.6: Training sample information for each real data set. Average training CE SNR is the average SNR of the CEs selected for training for a given data set. The number of windows the training noise sample for each data set was split into is also listed.

Data Set		Average Cond. CE SNR	Number Cond. Noise Windows
fMCI-PC	1	8.23	91
Somatic fMCI	1	1.80	328
	2	1.22	159
	3	1.51	152
Dendritic Spine fMCI	1	2.51	162
	2	3.84	124
	3	2.95	183
	4	8.87	116
	5	2.34	211

5 dendritic spine fMCI data sets were used for this experiment. Note that no ground-truth was available for this data.

Training for Real Data The MMiCE detector was separately conditioned for each data set. Because the fluorescent trace, X_r, for a given ROI, r, is too short to serve as an adequate training sample, all of the fluorescent traces for a given data set were concatenated into a single time series for training. For each data set, 20 CEs with relatively high SNRs, and smooth shapes were then manually selected from each concatenated data and used to generate training templates s_l, for each data set l. An example of both a somatic and spine template is shown in Fig. 2.6(A). The concatenated data were also used to manually select 20 training noise samples for each data set. The length of each noise sample varied depending on the availability of long uninterrupted noise segments in the data. The average SNR of each data set's training CEs, as well as the number of windows the training noise could be subdivided into, is listed in Table 2.6. The SNR was calculated as shown in Eq 2.10. All data selected for training were omitted from the performance analysis.

Manual CE Identification in Real Data Six trained analysts, two of whom also collected the fMCI data, manually tagged CEs in the 3 somatic and 5 spine data sets. The average CE SNR for each data set and each analyst is shown in Table 2.7. The SNR was calculated as in Eq. 2.10, with σ^2_{Noise} calculated from the training noise sample from each corresponding data set. Agreement between analysts was defined as the percentage of CEs that all analysts who worked with the data set agreed are CEs and not noise. The agreement % for each analyst and each data set is listed in Table 2.7.

On average, the manually identified CEs in the somatic fMCI data sets had an SNR of 1.53. Similarly, the CEs identified in the spine data sets had an average SNR of 2.43. As mentioned previously, this very low SNR level is fairly typical of neural imaging data. Although some ROIs exhibited SNR levels as high as 30, this was very rare.

More importantly, the level of agreement between analysts was very low. For the somatic data sets, on average, analysts only agreed on 50.48% of the identified CEs. This means that on average 49.52% of CEs identified by any analyst in a given data set, were considered noise by at least one other analyst working with that data set. Similarly the average agreement rate between analysts for the spine data sets was only 31.52%. These results are even worse, meaning that 68.48%, or over two-thirds, of CEs identified by any analyst were considered noise by at least one other analyst. Examples of the kinds of CEs tagged by the analysts are shown in Fig. 2.10.

Overall, the analysts were not confident in their ability to accurately identify CEs at SNR levels of 3 and below, which constitute most of the events present in these data sets. The lack of agreement among analysts also shows that manual CE identification is unreliable and error-prone. As has also been concluded by Sasaki et al. [88], manually identified CEs are not a reliable estimate of the ground truth, especially at low SNRs. Furthermore, the process proved to be extremely laborious. On average, each analyst spent roughly 24 working hours identifying CEs in a single data set.

Table 2.7: Average SNRs of manually identified CEs for each analyst and each fMCI data set. Not all analysts identified CEs in all available data sets. The data sets not analyzed by a given analyst are shown as blanks. The agreement % among analysts for each data set is also shown. The agreement % is the fraction of CEs that all analysts who worked with the data set agree are CEs and not noise.

Data Set		SNR							Agreement %						
		Analyst							Analyst						
		1	2	3	4	5	6	Avg.	1	2	3	4	5	6	Avg.
Somatic fMCI	1	1.59	1.35	1.34	1.28	2.17	–	1.55	58.20	53.38	65.03	48.77	73.26	–	59.93
	2	2.20	1.45	1.53	2.01	1.93	–	1.83	71.79	50.79	67.11	74.93	58.81	–	64.69
	3	1.37	1.14	1.11	1.19	1.30	–	1.22	47.83	18.97	28.28	6.96	31.43	–	26.83
	Average	1.72	1.31	1.33	1.50	1.80	–	**1.53**	59.27	41.05	53.70	43.89	54.50	–	**50.48**
Dendritic Spine fMCI	1	1.88	1.79	1.78	–	–	2.47	1.98	19.05	13.46	25.47	–	–	38.07	24.01
	2	2.84	2.11	2.17	1.54	–	2.91	2.31	41.33	29.48	38.26	9.86	–	42.06	32.20
	3	2.23	1.77	2.11	1.35	–	2.20	1.93	34.01	22.83	29.45	6.95	–	42.43	27.15
	4	3.23	4.47	5.23	3.19	–	5.07	4.24	31.84	53.22	67.70	29.21	–	60.01	48.40
	5	1.42	1.49	1.48	–	–	2.07	1.62	22.73	33.21	17.63	–	–	48.84	30.60
	Average	2.32	2.33	2.55	2.03	–	2.94	**2.43**	29.79	30.44	35.70	15.34	–	46.28	**31.51**

MMiCE Detector Performance on Real Data The MMiCE detector was applied using the full noise covariance, and performed very well on real data. The detector could consistently identify CEs with a specified sensitivity level. This was not true for manually identified CEs, which were not consistent between traces or across data sets. Upon reviewing detection results with varying threshold sensitivity values, each analyst identified a threshold sensitivity value, a, that they considered to be optimal for each data set. The selected values ranged from 2 to 4 for the somatic fMCI data, and 1.7 to 3 for the spine fMCI data. This roughly corresponds to the thresholds for the best detection results at SNR 0.2 - 0.5 for the simulated data sets. The average SNR of the real data is 1.53 - 2.43 (Table 2.7), which, given the bias discussed in Sec. 2.3.3.3, is about equivalent to SNR 0.5 - 2 in the simulated data case. These thresholds should therefore achieve near optimal performance.

Figure 2.10 *Top* shows examples of detected CEs in both a low-SNR case (*Left*) and a high-SNR case (*Right*) for somatic fMCI data. Both data traces are from somatic fMCI data set 1, and both filter outputs are thresholded with $a = 3$. Similarly, Fig. 2.10 *Bottom* shows

the same for the spine fMCI data, with both traces from spine fMCI data set 4, and both filter outputs thresholded with $a = 3$. Note that the MMiCE detector was able to identify many CEs of various shapes (Fig. 2.10 *Bottom Right*), even though only a single template was used for each data set. Analysts indicated that they trust the filter output more than their own judgment in identifying low-SNR CEs.

Comparison with Manually Identified CEs MMiCE detected CEs were compared to the CEs identified manually. In this case there is no ground-truth to assess the detector's performance, therefore the analysts' judgment was used as the ground-truth. Detected CEs did not closely match the analysts' CEs. Although the detector could always accurately identify over 90% of any analyst's CEs for any given data set at a low enough threshold, this often resulted in an extremely high FP Rate, where again here a FP is defined relative to the CEs identified by the analyst. This discrepancy is largely due to the analysts' lack of consistency in CE selection. Even taking the CEs that all analysts agreed on did not result in a close match with the MMiCE detected CEs (TP $\simeq$ 70%, FP $\simeq$ 20%, data not shown). This is largely due to the bias, and within-analyst inconsistency (i.e. the analyst is not consistent with herself) of the analysts. This is further discussed in Sec. 2.3.3.4. Furthermore, when analysts were shown the MMiCE detected CEs compared with their own identified CEs, they were more likely to fault themselves with an omission or false positive than to attribute an error to the detector. Lastly, the MMiCE detector greatly reduced detection time. The computation time for CE detection was tested using Matlab and is 0.0047-s per every second of recording per ROI (0.0047-s/ROI-s). This amounts to 75.0 $\pm$ 47.6-s on average per data set, depending linearly on the length of the recordings, and the number of ROIs in the data set. The same data sets took about 24 hours to tag manually.

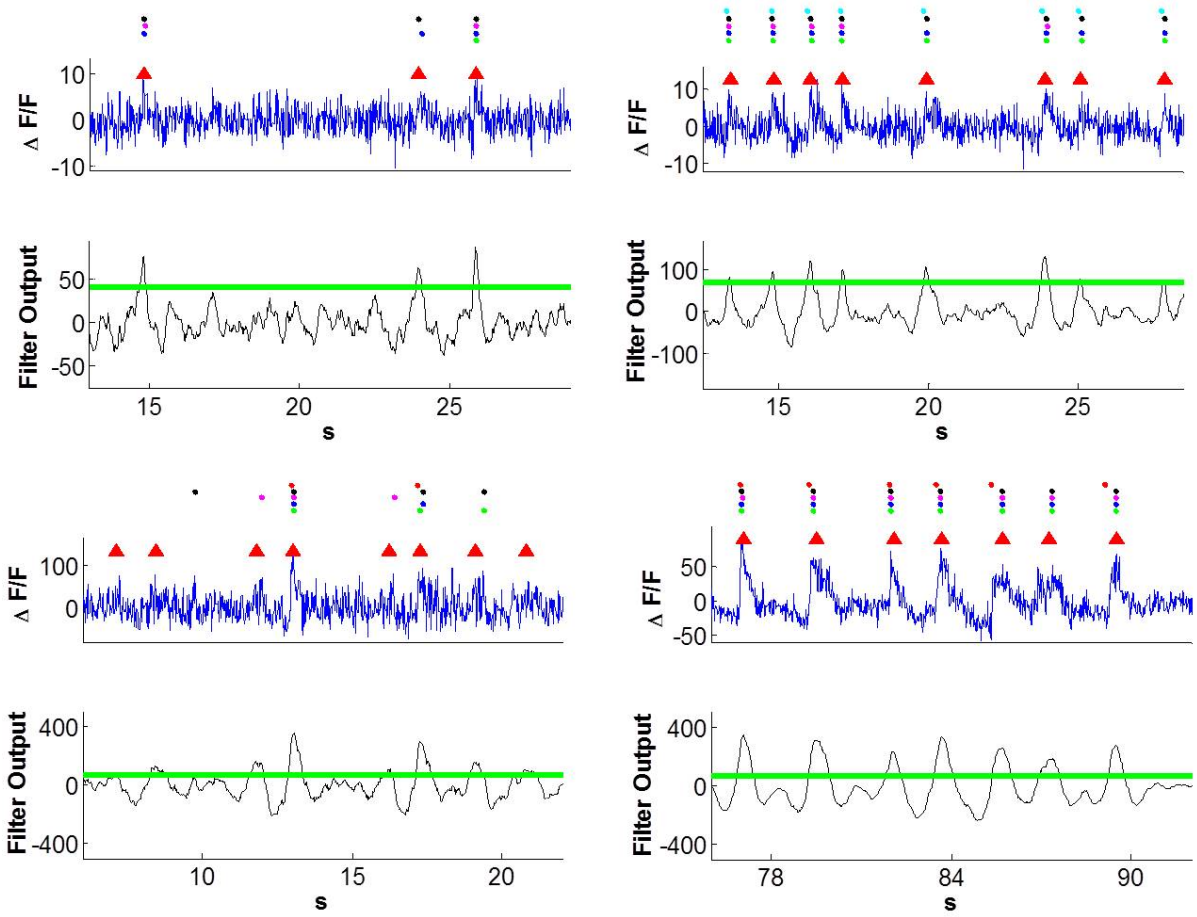

Figure 2.10: Four examples of CEs detected from real data. Each example consists of a *top* trace showing the raw Δ F / F data with MMiCE detected CE peaks marked by red triangles, and a *bottom* trace showing the filter output and the detection threshold used. The circles, going from bottom to top, represent spikes identified by analysts 1-6 respectively. The top two examples are from somatic fMCI data set 1 in both a low-SNR case (*Left*) and a high-SNR case (*Right*). The bottom two examples are from spine fMCI data set 4 in both a low-SNR case (*Left*) and a high-SNR case (*Right*). All four filter outputs shown here were thresholded with $a = 3$. Note that in the spine low-SNR case (*Bottom Right*) the MMiCE detector was able to identify many different CE shapes even though only a single training template was used. This is also the case where analysts did not agree on most of the spikes. In the high-SNR cases (*Right*), the MMiCE detector agrees more closely with a majority of the analysts, although even here the analysts are not unanimous about all of the spikes.

2.3.3.4 Discussion

WGN Assumption The MMiCE detector was tested both under a WGN approximation, and using the full noise covariance. In general, the noise in neurological data is not white, as

a large portion of it is biological noise coming from neighboring neurons [10]. Our results in simulated data, however, indicated that using a WGN approximation is sufficient for precise and accurate CE detection, especially at SNR ≥ 1. Once the SNR falls below 1, a WGN approximation is no longer sufficient and the full noise covariance should be used.

Reliability of Manual CE Identification As shown in Sec.2.3.3.3, the level of agreement between analysts was very low (50.48% for the somatic data, and 31.52% for the dendritic spine data). As all analysts were thoroughly trained in identifying CEs, this discrepancy is not due to any single analyst simply mis-identifying a majority of events. For the most part, the disagreement occurred in low-SNR traces, where the signal was difficult to identify accurately. There was also disagreement, especially in the spine data sets, due to overlapping CEs and variations in CE shape. While some analysts would identify wide and oddly shaped CEs as more than one event, others considered them a single CE. This discrepancy is the reason why the spine data has a much lower agreement rate, while maintaining a higher average SNR than the somatic data. As SNR increased and overlapping CEs exhibited two or more clearly identifiable peaks, this discrepancy disappeared.

The analysts also proved to be inconsistent in CE identification. Although in one trace a given analyst may identify a CE of SNR 1.5, in another trace they may fail to identify any CEs below SNR 3. This is especially dependent on the number of higher SNR CEs in the same trace. If there are CEs of SNR 5 and above in a trace, then spikes of SNR 1.5 will be omitted by analysts. Conversely if the trace is mostly noise, or only has some CEs of SNR ≤ 3, then CEs of SNR 1.5 are very likely to be identified. This bias was true for all analysts. Overall, manual CE identification is unreliable and error-prone.

Effects of Differing Training Samples Although only 20 samples of both CEs and noise segments needed to be identified for training on a given data set, some may still find this

process too laborious. It may therefore be advantageous to re-use training data samples on several data sets. The MMiCE detector's efficacy was therefore tested with the training samples and test data coming from different data sets, and the results were qualitatively compared.

When training samples for different somatic fMCI data sets were applied to the somatic fMCI data, the detection results were virtually unchanged. Although the actual amplitude of the filter output changed, the relative peaks in the filter output stayed consistent. A slight dip in performance was observed when using training samples from somatic fMCI data set 2. This training sample contained the lowest SNR CE samples (Avg. SNR = 1.22, Table 2.6) of all of the data sets. The slight dip in performance, and low SNR of training CEs may indicate that the training template s from this data set was poorly determined.

Similarly, if spine fMCI training data was varied between the spine fMCI data, the detection results remained roughly the same. No dip in performance was observed for any training sample. The filter output peaks were more pronounced when using the training CEs from spine fMCI data set 4 (Avg. SNR = 8.87, Table 2.6), but this did not result in markedly improved detection.

Detection performance did worsen when spine fMCI training data was used with somatic fMCI data, but not vice versa. This is largely attributed to the fact that the spine training CEs were wider than the somatic CEs. This caused the filtered CEs to be wider and lower in amplitude in the case of somatic fMCI test data, which in turn may have negatively affected detection. When somatic fMCI training spikes are used on spine fMCI data, the skinnier CE template caused the filtered CEs the be thinner and higher in amplitude, which did not significantly affect spike detection.

In general, re-using training samples from one data set to detect spikes in other data sets did not negatively affect detection performance. Performance only degraded if the training CEs

are of low quality and very low SNR. Although it is possible to use training samples from one data type (such as somatic fMCI) on other data types (such as dendritic spine fMCI), it is not recommended for best performance.

Note, however, that the imaging set-up for the data collected here was consistent. Furthermore, one experimentalist collected all somatic fMCI data (Analyst 5 from Table 2.7) and another experimentalist collected all dendritic spine fMCI data (Analyst 6 from Table 2.7). Therefore, performance may degrade if training samples from data collected by a given experimentalist are used on data collected by a different experimentalist whose procedure may differ. Similarly, performance may degrade if training samples are used on data with a different imaging set-up or a different sampling frequency. For best performance it is recommended that training samples are only re-reused on data of the same type, collected by the same experimentalist / under the same protocol, and using the same imaging set-up. Otherwise, performance may be affected. Training samples from the fMCI-PC data were not re-used because this data was taken at a sampling frequency of 10Hz, where as the other real fMCI data was taken at a sampling frequency of 50Hz. The training samples were therefore not compatible.

Limitations Like many pre-existing methods [12, 40, 47, 66, 81, 88], the MMiCE detector does not focus on image processing, and requires ROIs to be pre-determined and pre-processed into fluorescence intensity traces before being fed into the algorithm. As the number of neurons that can be simultaneously imaged continues to increase, this may prove to be a major limitation of this approach. Not only is manual ROI selection time consuming, but it may also be subject to bias. Our previous work addresses this issue by developing a stand alone algorithm for automatically extracting ROIs from somatic fMCI data [70]. Although not applied here, the MMiCE detector can be easily paired with either this method, or other algorithms for ROI extraction such as the ones presented in Mukamel et al. 2009 [68]

or Cossart et al. 2003 [12]. Note, however, that the automated ROI extraction tools listed here are developed for somatic fMCI data. To the best of our knowledge, manual ROI selection is therefore still necessary for dendritic spine fMCI data.

Another limitation of this approach is the need for manual inspection of the data to obtain training samples. Using a fully automated algorithm that relies on general templates from a library of pre-constructed templates or on canonical filters such as wavelets, would remove the need for human involvement. However, this automation step would come at the expense of lowering detection performance. While CEs across different datasets tend to exhibit stereotypical features, there are sufficient dataset-to-dataset differences that may not be optimally accommodated with general templates. Furthermore, the general templates may themselves be biased to the dataset that they were constructed from. This bias may be mitigated with the use of wavelets. However, since they have to be zero-mean functions, wavelets cannot be monophasic and are therefore suboptimally matched to the shape of CEs. Therefore, performance is expected to degrade significantly if wavelets are used to replace the data-driven CE templates presented in this study.

2.3.3.5 Conclusion

In this study we presented a Matched filter for Multi-unit Calcium Event (MMiCE) detection that is designed to identify CEs in low-SNR environments. The algorithm is completely data driven, making it applicable to a wide range of data types and testing environments. The MMiCE detector was tested on simulated somatic fMCI data, and performed exceptionally well. The simulated data was constructed from noise segments and CEs captured from real neuronal somatic fMCI recordings and varied from SNR 0.2 to SNR 14. Even at SNR 0.2 the MMiCE detector reached a TP Rate of 90.46 (3.11)% and a FP Rate of 12.81 (3.47)% when applied in its simplest form, assuming WGN statistics. The performance was even better when the full noise covariance was used in training, with a TP Rate of 98.27 (1.38)% and a FP

Rate of only 6.59 (2.56)% at SNR 0.2. The detector reached perfect performance (TP Rate = 100 (0.00)%, FP Rate = 0 (0.00)%) at SNR 2 and above, even under WGN assumptions. The MMiCE detector's performance was also tested on simultaneously recorded somatic fMCI and patch-clamp data. The patch-clamp recording provided a ground-truth to quantitatively assess the MMiCE detector's performance on real somatic fMCI data. Under the WGN assumption the MMiCE detector reached a TP Rate of 95.83 (5.86)% and a FP Rate of 2.13 (4.28)%, which constitutes 3 errors. When the full noise covariance was used the results improved to TP = 100.00 (0.00)% and FP = 2.04 (4.10)%, which constitutes only 1 error. This high performance level was on par with and in some cases exceeded that shown by existing methods [47, 88, 66, 119, 27]. Although the MMiCE detector did not attempt to extract AP firing rates from the detected CEs, it is worth noting that it achieved similar performance while being tested at SNR levels as low as 0.2, which are well below those used in previous studies (SNR $\simeq$ 5 - 10).

Lastly, the MMiCE detector's performance was qualitatively presented for experimentally recorded low-SNR somatic fMCI and dendritic spine fMCI data. The detector performed very well, able to consistently detect CEs with a specific sensitivity level. Furthermore, the detector was able to detect CEs of varying shapes and widths, even though only a single training template was used for detection. Analysts who also identified CEs in the data sets proved to be unreliable. The level of agreement between analysts in identifying CEs was only 50.48% for the somatic fMCI data sets, and only 31.51% for the spine fMCI data sets. Analysts indicated that they trust the MMiCE detector's output over their own judgment, especially for CEs with SNR $\leq$ 3. Furthermore the MMiCE detector reduced detection time from 24 hours per data set manually to 75-s per data set on average (0.0047-s/ROI-s).

In order to accommodate the varying signal amplitudes and SNR levels found in different ROIs of a given imaging data set, the MMiCE detector applied an innovative thresholding approach. The threshold was determined independently for each ROI by calculating

the detection filter output's median and median standard deviation. These quantities effectively estimate the mean and standard deviation of the filtered noise, respectively. The filter threshold is then set as a multiple, a, of median standard deviations above the median, where a is a thresholding sensitivity parameter set at the onset of detection. Optimal thresholding sensitivity levels, a, set by the analysts ranged from 1.7 to 3 for spine and 2 to 4 for somatic data. Optimal thresholds for the MMiCE detector using the full noise covariance on simulated data with a similar SNR range (SNR 0.2 - SNR 1) were $a = 5.00 \pm 1.37$. The optimal thresholding sensitivity level for the fMCI-PC data ranged from 3 to 4. This thresholding approach allows the MMiCE detector to accurately and consistently extract CEs with a specified sensitivity level even across wide variations in signal and noise amplitudes.

The MMiCE detector was also shown to be very robust. Detection retains fidelity even when training samples come from a data set that is not the one being tested. It is even possible to share training samples across different data types, such as between somatic fMCI and dendritic spine fMCI data. However, performance does begin degrading at this point, therefore it is not recommended.

Overall, these results indicate the MMiCE detector is applicable for widespread use in detecting low-SNR neuronal CEs from imaging data.

Acknowledgments I would like to extend a special thanks to N. Koe, M. Doty, and K. Scannell for their help in manual CE identification.

2.3.4 Unsupervised Automation Approaches

In general, automated approaches to spike detection can be supervised or unsupervised. Supervised detection assumes that measurements of both spikes and noise are available before the detector is implemented. The matched-filter approach presented up to this point

falls under this category as both spikes and noise segments are identified in the training process, before the filter is applied. The thresholding process is also supervised, as the user is given the option of selecting a desired threshold sensitivity value.

It may be desirable to further automate these processes. Although further automation goes hand in hand with decreased performance, large data sets, or data sets where representations of spikes are not readily available may benefit from further automation. In the work presented below we go into some detail about possible automation approaches that can be applied to the matched filter discussed throughout this chapter. More specifically, we will discuss automating threshold setting, as well as automating the spike template construction via the use of continuous wavelet transforms (CWTs). Although CWTs could be used with various spikes including DEs and CEs, their shape is most applicable to APs, hence this modality is where the automation step is explored.

2.3.4.1 Automating Sensitivity Setting

As seen in Sec. 2.3.1- 2.3.3, different analysts often disagree on an appropriate threshold sensitivity setting, and tend to be poor estimators for the ground-truth. One of the best ways to minimize bias in detection is therefore to completely automate threshold setting. Assuming spikes are outliers within a Gaussian noise distribution, we can define an outlier threshold as

$$\gamma = M + \sigma_M \sqrt{2 \ln T} \tag{2.11}$$

where T is the number of time samples spanning a data trace X, and σ_M is derived in Eq. 2.4 [71]. This threshold is essentially an upper bound on the noise distribution, therefore anything above it should belong to a different distribution, i.e. the signal.

For the extracellular AP data presented in Sec. 2.3.1 this gives $\gamma' = 4.88$ and is roughly equivalent to $a = 6.0$, falling somewhere between the sensitivity levels of Analysts 1 and 2.

For the fMCI-PC data trace presented in Sec. 2.3.3 this gives $\gamma = 76.71$. This is equivalent to a threshold sensitivity setting of $a = 3.9$, which falls within the range of optimal threshold sensitivity settings ($a = 3 - 4$) when MMiCE is applied under the WGN assumption (Sec. 2.3.3.3, Fig. 2.9). It is higher than the optimal threshold when the full noise covariance is used, which was $a = 2$. For the 5 dendritic spine fMCI data sets presented in the same sections (Sec. 2.3.3) this gives an average threshold of $\gamma = 51.48 \pm 27.49$, which is equivalent to a threshold sensitivity of $a = 4.16 \pm 0.08$. Similarly, the 3 somatic fMCI data sets had an average optimal threshold of $\gamma = 69.96 \pm 28.71$, corresponding to a threshold sensitivity of $a = 4.03 \pm 0.06$. The variability in γ is expected as it is intended to differ significantly based on signal amplitude. The threshold sensitivity, a, on the other hand should remain relatively constant throughout the data sets, which is also reflected here. This threshold sensitivity is slightly above that set by analysts (1.7 to 3 for dendritic spine, and 2 to 4 for somatic data), but corresponds well with the optimal threshold determined for the SNR 0.5 simulated data set ($a = 4.00 \pm 0.50$).

Overall, this approach seems to match optimal or near optimal thresholds presented in this work, and is therefore a good candidate for implementing a more automated approach. However, the algorithm breaks down if spikes begin to become prevalent enough in the data that they can no longer be considered outliers. This is very rare in extracellular recordings, but seen in some cells with fMCI, and very common in cardiomyocyte recordings. The automation step is therefore most appropriate for extracellular recordings, and should not be used with cardiomyocyte data.

2.3.4.2 Automating Spike Template Generation

This section presents a rational design approach to building a multi-sensor extracellular AP detector. The proposed method is rooted in statistical detection theory which is modified to account for scenarios when the signals of interest are not known. Based on biophysical constraints on the duration and shape of APs, we argue that the SNR of extracellular recordings can be significantly improved using an appropriate continuous wavelet basis.

Background On Unsupervised Multi-sensor Extracellular AP Detection In unsupervised (blind) detection, noise and AP templates are generally unknown. However, a similar problem is commonly encountered in approximation theory, where blind signal/noise separation is effectively accomplished in a suitable sparse representation basis [61]. Motivated by these ideas, Nenadic and Burdick [71] applied Bayesian decision theory in a continuous wavelet transform (CWT) domain to develop a fully unsupervised AP detection algorithm. An extension of this work utilizes probabilistic models to identify APs as outliers in a noise distribution [6]. However, both of these methods are only suitable for single-sensor data, and to the best of our knowledge, there are no methods for unsupervised AP detection in multi-sensor recordings.

Extracellular recordings often generate measurements that contain no visually discernible APs. From this so-called noise-only data, the noise statistics can be estimated and utilized in AP detection. Falling between the extreme cases of supervised and unsupervised detection, this approach to AP detection can be referred to as semi-supervised, since it utilizes the noise statistics but makes no use of information about the signal. In the context of tetrode recordings, power detection [15] and a prewhitening transform [85] are good representatives of this approach. The former method calculates the power of data over a short time segment and compares it to that of background noise. Since the power of APs is expected to be above the baseline, a suitable threshold can be defined that separates APs from noise [15, 31].

The prewhitening approach, on the other hand, uses the spatial covariance matrix of noise to decorrelate noise samples across channels and normalize their variance to 1. Detection thresholds are then set for each channel in the prewhitened space [85].

The approach we present here will use the semi-supervised paradigm with the matched filter presented previously. Namely, the signal template s (Eq. 2.1) will be estimated using the CWT, as opposed to from the data; Σ (Eq. 2.1), representing the noise statistics, will, however, still be estimated from the data. Our prior work [71, 6] provides extensive arguments for using the CWT of the biorthogonal class for AP representation. Briefly, the shape of these wavelet functions is reminiscent of the predominantly biphasic shape of many APs [39]. Furthermore, since APs are highly localized in time with characteristic durations of $\sim$1 ms [39, 54], an extremely large set of CWT scales can be significantly reduced. Finally, the translation invariance of the CWT ensures that the representation of an AP does not depend on its relative position within the time series. These last two properties of the CWT make it more suitable to AP detection than the commonly used discrete wavelet transform with dyadic scales and translations [121, 76].

Methods To modify the matched filter in Eq. 2.1, we assume that the spike, $\underline{s}$, can be well estimated by discrete samples of compactly-supported wavelet functions, $\underline{\psi_{\alpha,\beta}} \in \mathbb{R}^{1\times N}$, with scale $\alpha > 0$ and translation β, where the same wavelet is applied across all channels c. The spike representation then takes on the form

$$\underline{s} = \underline{\mathbf{\Psi}_{\alpha,\beta}} = \begin{bmatrix} \underline{\psi_{\alpha,\beta}} \\ \underline{\psi_{\alpha,\beta}} \\ \vdots \\ \underline{\psi_{\alpha,\beta}} \end{bmatrix} \in \mathbb{R}^{C\times N} \tag{2.12}$$

and we can rewrite Eq. 2.1 as

$$S_{\alpha,\beta}(x) = \mathbf{\Psi}_{\alpha,\beta}\Sigma^{-1}x^T \quad , \quad \begin{array}{lll} H_1 & \text{if} & S_{\alpha,\beta}(x) > \gamma \\ H_0 & \text{if} & S_{\alpha,\beta}(x) < \gamma \end{array} \tag{2.13}$$

where $\mathbf{\Psi}_{\alpha,\beta}$ is the row-vector form of $\underline{\mathbf{\Psi}_{\alpha,\beta}}$. Since β is a translation measure, we can also think of $S_{\alpha,\beta}(x)$ as a single-scale test statistic $S_\alpha(X) \in \mathbb{R}^{1\times T}$, resulting from the template $\mathbf{\Psi}_{\alpha,\beta}\Sigma^{-1}$ being convolved over the entire data trace X (Sec. 2.2.1.3), and summed across sensors, for a given wavelet with scale α.

The data used to investigate this approach is the same data as in Sec. 2.3.1, detailing the matched-filter approach for extracellular action potential detection. The data is available online [78], and additional details about the data collection procedure beyond that provided in Sec. 2.3.1 are described in Pouzat et al. 2002 [79]. To objectively measure SNR, 20 APs and 25 noise segments (duration between 10 and 60 ms), were randomly selected and manually delineated in the time series data. Extracting longer noise segments was not possible due to the high firing rates of the neurons. The noise samples were used to calculate Σ, whereas the APs was used for validation.

Based on a typical AP duration, we assumed a signal length of 2 ms ($N = 31$). For each noise segment and each sensor pair $\{i = 1, 2, 3, 4;\ j \geq i\}$, cross- and auto-covariance sequences, $r_{i,j}(k)$, were calculated at lags $k \in \{-30, -29, \cdots, 30\}$. These sequences were then averaged over the 25 noise segments to obtain stable estimates. Note that unlike auto-covariance, cross-covariance is not guaranteed to be an even function, although the values of $r_{i,j}(k)$ and $r_{i,j}(-k)$ were very close. Therefore, the cross-covariance sequences were averaged over the positive and negative lags, i.e. $\bar{r}_{i,j}(k) = 0.5\,[r_{i,j}(k) + r_{i,j}(-k)]$, $k \in [0, 30]$, resulting in a total of 310 (31×10) parameters. Finally, the p-values of $\bar{r}_{i,j}(k)$ were estimated by a Monte-Carlo simulation, and those $\bar{r}_{i,j}(k)$ deemed statistically insignificant ($p \geq 0.05$) were set to 0. This

further reduced the number of parameters to 269. The averaged and reduced sequences were then used to generate Σ.

Biorthogonal wavelets (bior1.5 in MATLAB $^{\text{TM}}$Wavelet Toolbox.) were chosen for analysis due to their AP-like shape. The scale α was chosen so that the duration of the dominant two phases of the wavelet matches that of an AP. For most APs, this falls between 0.5 and 2 ms [39, 54], thus 16 scales were chosen to cover the $[0.5, 2.0]$ ms range in 0.1-ms increments.

For each selected AP, $u_i \in \mathbb{R}^{4\times 31}$ ($i \in \{1, 2, \cdots, 20\}$), and noise segment, $n_j \in \mathbb{R}^{4\times d_v}$ ($j \in \{1, 2, \cdots, 20\}$; d_v-variable), the SNR at sensor c was defined as:

$$\text{SNR}_c(u) = \underset{j}{\text{median}} \left\{ \frac{\|u_i\|_\infty}{\|n_j\|_\infty} \right\} \tag{2.14}$$

Note that since nearly all detectors are threshold based, the SNR is more appropriately defined using the L^∞-norm than the commonly used L^2-norm. Apart from scaling the SNR values, the use of the L^2-norm did not affect the results presented below.

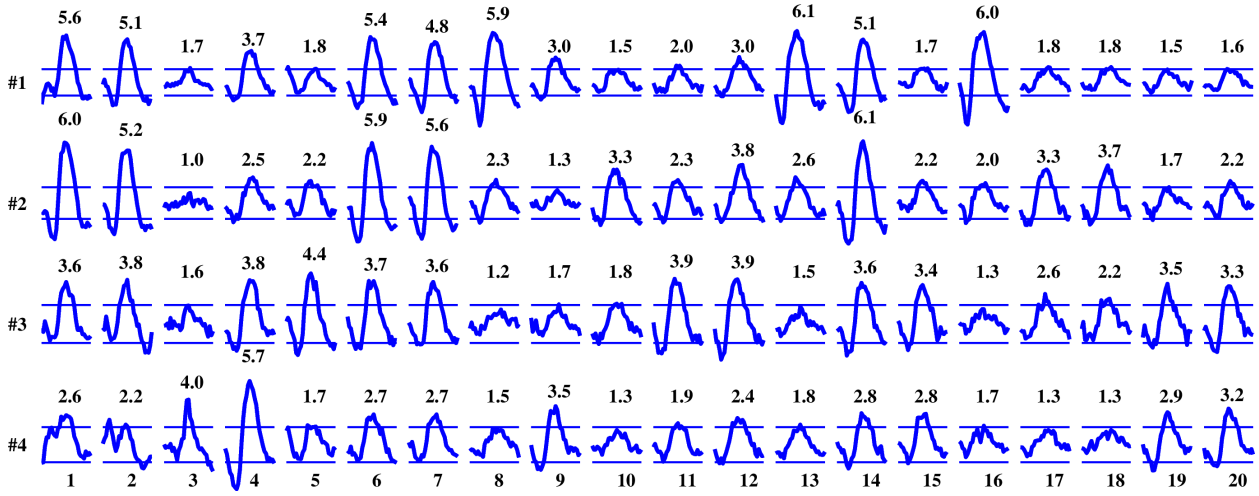

Figure 2.11: A sample of 20 APs recorded with a tetrode (APs labeled 1-20, sensors labeled 1-4). Each trace is 2 ms long. The horizontal lines mark $\pm 5\sigma$ bounds, and the number next to each waveform represents its median SNR.

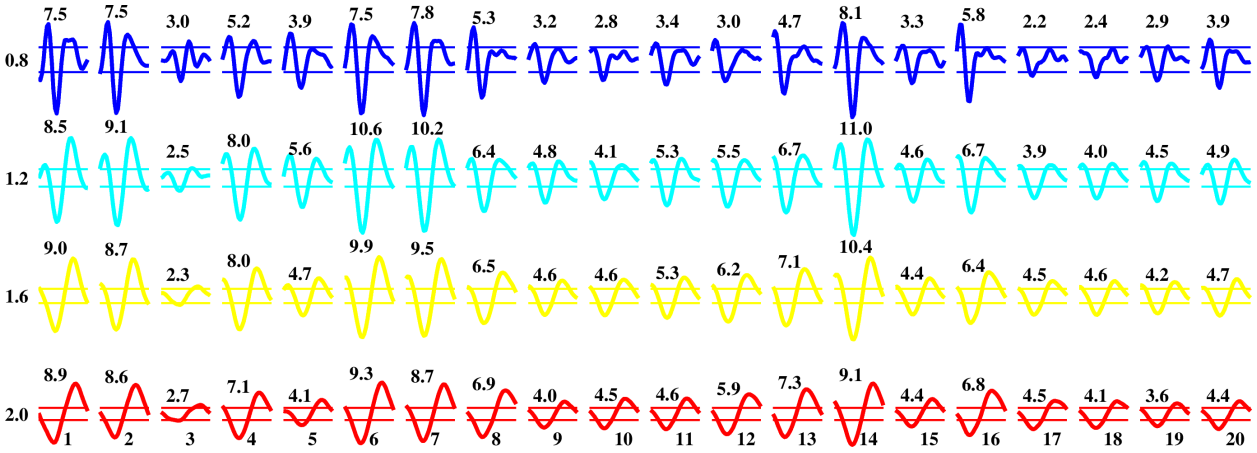

Figure 2.12: Test statistic of 20 APs at 4 wavelet scales corresponding to the durations 0.8, 1.4, 1.6 and 2.0 ms. Each trace is 2 ms long. The horizontal lines mark $\pm 5\sigma$ bounds estimated from the test statistic of noise, and the number above each waveform represents its median SNR.

Results Fig. 2.11 shows the APs selected for validation as well as the distribution of SNRs across the APs and sensors. Visual inspection confirms the diversity of the sample, as there appear to be at least 6 distinct classes of APs. The average SNR ranged from $\sim$2 (APs 3 and 10) to $\sim$4.5 (APs 1 and 6). Also shown are the noise bounds estimated by averaging the noise standard deviation over the 25 noise segments.

Fig. 2.12 shows a similar plot for the data processed using the wavelet-approximated generalized matched filter (AGMF) (Eq. 2.13). In this case, the SNR was not calculated across sensors, c (Eq. 2.14), since the matched filter sums across all of the sensors; it was instead calculated as a function of the wavelet scale α (which therefore replaced c in Eq. 2.14). In the interest of space, Fig. 2.12 shows the EAP test statistic at 4 representative scales. Even from this reduced set of scales, it is apparent that substantial SNR improvements are achieved by the AGMF.

To formally ascertain the degree of SNR improvement, we performed statistical tests on the SNR of the 20 APs. The maximum SNR per sensor was calculated for each of the 20 APs in the original data. Similarly, the maximum SNR per scale was calculated for each AP in the AGMF-processed data (Fig. 2.13). A sign test showed that the SNRs of the processed

APs were significantly superior to those of the original data ($p < 0.00004$), with a median improvement of 38%. A similar test was performed to compare the maximum SNR per sensor to the SNRs at the 16 individual scales (i.e. without taking the maximum over scales). It was found that the SNRs at 10 scales (1.1–2.0 ms) were superior to the maximum SNR based on the original data. These results indicate that even if the analysis is performed at a single wavelet scale, the SNR improvements with respect to the original data are significant (Fig. 2.13).

Fig. 2.13 also shows the distribution of SNRs across APs achieved with the power method [15, 31] and spatial prewhitening [85]. The power method was implemented by calculating the root-mean-square value of the data at each sensor in a 2-ms sliding window. The SNR on a per sensor basis was then calculated using(Eq. 2.14). Finally, the spatial prewhitening method was implemented by averaging the spatial covariance matrix over 25 noise segments. The resulting matrix was then spectrally decomposed, and the prewhitening matrix was calculated [116] and used to premultiply the tetrode data. As expected given Fig. 2.13, a sign test showed that the SNR of the power method was significantly superior to the original SNR ($p < 0.0004$), with a median improvement of 22%. The SNR of the prewhitening method, however, was inferior to the original SNR ($p < 2 \times 10^{-6}$), with a median loss of 10%. In addition, both the power and prewhitening methods were significantly inferior to the AGMF method, with $p < 0.01$ and $p < 0.00004$, respectively, and a median loss of 18% and 35%, respectively. The results were also compared to two popular unsupervised techniques: principal and independent component analysis (PCA and ICA, respectively). PCA yielded a median SNR improvement of 6% with respect to the original data, however, this gain was not statistically significant (Fig. 2.13). ICA, on the other hand, resulted in a significant loss (20%) of SNR with respect to the original data (Fig. 2.13).

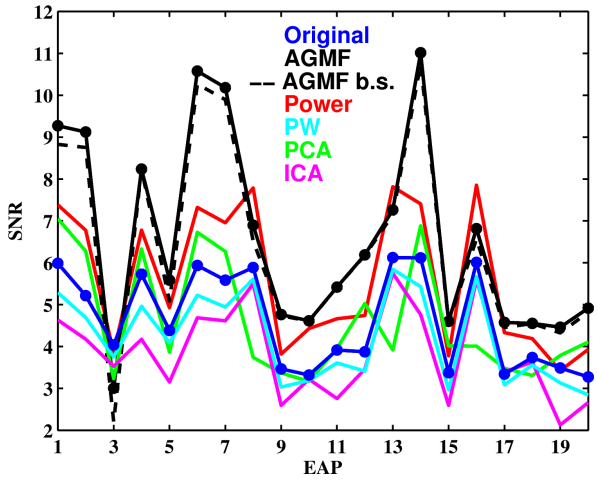

Figure 2.13: Maximum (per sensor) SNR of the original, power, and spatially prewhitened (PW) data. Maximum (per scale) and best scale (b.s.) SNR of the AGMF-processed data. Maximum (per component) SNR of the PCA- and ICA-processed data.

Discussion and Conclusion By combining statistical detection theory with continuous wavelet representation, we derived an approximation of the generalized matched filter suitable for semi-supervised AP detection problems, where the noise samples are available, but the signal samples (spikes or APs) are not. The method outperforms other multi-sensor AP detection approaches by yielding a significantly higher SNR when tested on representative AP and noise samples extracted from experimental data. A notable exception is AP 3, which is characterized by weak signals at 3 sensors (Fig. 2.11). This is hardly surprising as the test statistic, $S_{\alpha,\beta}(x)$, combines data across sensors, which in this case "dilutes" the SNR. Our work also shows that the use of *ad hoc* multi-sensor data processing methods may not only fail to improve the SNR, but may actually lower it. Future studies will focus on a formal testing of the AGMF method for the detection of APs under different sensitivity-specificity trade-offs and in the presence of APs with temporal overlap. Finally, we will also pursue the development of a fully unsupervised multi-sensor AP detection method.

2.4 Conclusion

In this chapter we've gone over three implementations of a supervised matched filter for spike detection from time series data. Each application concentrates on a different data type, starting from multi-sensor extracellular recordings, to VSD imaged cardiomyocytes, and finally calcium imaged neuronal somas and dendritic spines. The chapter ends with some exploration of more automated approaches.

2.4.1 Multi-sensor Extracellular Action Potential Detection

The first matched filter application tested the detector on tetrode recordings from a locust antenna lobe. Because no ground truth was available, detector performance was assessed by comparing detection results to APs tagged manually by 3 trained analysts. To reduce complexity and avoid poorly estimating the noise statistics, the filter assumed that the data noise was spatially white. The detector performed with an average TP rate of 85% and an average FP rate of 17%, as determined by ROC curves comparing performance to the APs tagged by the trained analysts.

The optimal thresholding sensitivity values, a, determined by the ROC curves, varied between the different analysts. The most selective analysts, Analyst 1, tagged APs with an average SNR of 28 and had an optimal thresholding sensitivity of $a = 7$. Analyst 2 tagged APs with an average SNR of 20 and had an optimal thresholding sensitivity of $a = 5.5$. The most liberal analysts, Analyst 3 tagged APs with an average SNR of 7 and had an optimal thresholding sensitivity of $a = 3.5$.

In general the agreement between analysts was poor (as low as 39%), which therefore implies that analyst's tagged spikes are not a good estimate of the ground-truth. The detectors performance is therefore likely underestimated in this study. Nevertheless, the detection

efficacy was still comparable to other methods, and the detector's performance presents it as a great candidate for wide-spread use.

2.4.2 MaD

The second application of the matched filter concentrated on detecting DEs from VSD imaged cardiomyocytes. Once again no ground-truth was available, so detection efficacy was tested against DEs manually identified from the data by trained analysts. DEs that were unanimous among all analysts were considered the ground-truth. The MaD detector was tested primarily using the full noise covariance.

Three different data cases were considered, isoproterenol treated hiPS-CMs, propranolol treated hiPS-CMS, and controls hiPS-CMs. The MaD detector was fist tested on typical data with an SNR range from 5 to 70. In the control case (Control-1, SNR = 6), the detector reached a TP rate of 98% with a 95% confidence interval of [92, 100]%, and a FP rate of 2 [0.00, 5]%. In the isoproterenol case (Iso-1, SNR = 70) the detector reached a TP rate of 100 [100, 100]%, and a FP rate of 1 [0, 2]%. Lastly, in the propranolol case (Pro-1, SNR = 31) the detector reached a TP rate of 99 [96, 100]%, and a FP rate of 1 [0, 4]%. In order to also test the detector in a more hostile, low-SNR setting, two low-SNR test cases were also considered, a control and isoproterenol treated hiPS-CMs. In the low-SNR control case (Control-2, SNR = 3) the detector reached a TP rate of 72 [62, 82]%, and a FP rate of 17 [8, 26]%, whereas in the low-SNR isoproterenol case (Iso-2, SNR = 1) the detector reached a TP rate of 66 [50, 82]%, and a FP rate of 14 [0, 27]%. Clearly, the performance decreased for these low-SNR data sets. We therefore conclude that the MaD detector performs best when applied to VSD imaged cardiomyocyte data of SNR $\geq$ 5. However, like in the extracellular recording case, the low-SNR data sets suffered from poor analyst agreement (as low as 36%) making the ground-truth estimate unreliable. It is possible that performance would change

if a more reliable estimate of the ground truth was available. Overall, the MaD detector is a useful tool for cardiomyocyte DE detection from VSD imaging data.

2.4.3 MMiCE

The final matched filter application was used on calcium imaged neuronal somas and dendritic spines. The MMiCE detector was designed to identify CEs in low-SNR environments, and tested on both simulated and experimentally recorded fMCI data. This time ground-truth data was available and the detector performed exceptionally well in both cases.

The simulated data was constructed from noise segments and CEs captured from real neuronal somatic fMCI recordings and varied from SNR 0.2 to SNR 14. MMiCE reached perfect performance at SNR 2 and above. Even at SNR 0.2 the MMiCE detector reached a TP Rate of 90 [87, 94]% and a FP Rate of 13 [9, 16]% under WGN statistics, and a TP Rate of 98 [97, 100]% and a FP Rate of only 7 [4, 9]% using the full noise covariance. The MMiCE detector's performance was also tested on simultaneously recorded somatic fMCI and patch-clamp data. The patch-clamp recording provided a ground-truth to quantitatively assess the MMiCE detector's performance on experimentally recorded somatic fMCI data. Under the WGN assumption the MMiCE detector reached a TP Rate of 96 [90, 100]% and a FP Rate of 2 [0, 6]%, which constitutes 3 errors. When the full noise covariance was used the results improved to TP = 100 [100, 100]% and FP = 2 [0, 6]%, which constitutes only 1 error. This high performance level was on par with and in some cases exceeded that shown by existing methods [47, 88, 66, 119, 27], and did so while being tested at SNR levels as low as 0.2, well below those used in previous studies (SNR $\simeq$ 5 - 10).

Lastly, the MMiCE detector's performance was qualitatively presented for experimentally recorded low-SNR somatic fMCI and dendritic spine fMCI data. The detector performed very well, able to consistently detect CEs with a specific sensitivity level. Analysts who

also identified CEs in the data sets proved to be unreliable and agreed on only 50% for the somatic fMCI CEs, and only 32% for the spine fMCI CEs. Analysts indicated that they trust the MMiCE detector's output over their own judgment, especially for CEs with SNR ≤ 3.

The MMiCE detector was also shown to be very robust. Detection retained fidelity even when training samples come from a data set that was not the one being tested. It is even possible to share training samples across different data types, such as between somatic fMCI and dendritic spine fMCI data. However, performance does begin degrading at this point, therefore it is not recommended.

Overall, these results indicate the MMiCE detector is applicable for widespread use in detecting low-SNR neuronal CEs from imaging data.

2.4.4 Further Automation

Although a fully composed detector using even more automated approaches was not designed, we did explore some further automation methods that could be used with the matched filter presented in this Chapter.

The first proposed automation step is unsupervised thresholding. Here we defined an outlier threshold, assuming that spikes are outliers within a Gaussian noise distribution. This assumptions seems to work very well, as thresholds calculated in this way provided near optimal results for both the extracellular tetrode data as well as the calcium imaging data tested in Sec. 2.3.1 and Sec. 2.3.3, respectively. This approach is however not appropriate for the cardiomyocyte data presented in Sec. 2.3.2. In this case the APs appear too often to be outliers, so the main assumption behind this automated thresholding approach is broken and the resulting threshold is too high.

The second proposed automation step is to replace the signal template with a CWT. Although CWTs could be used with various spikes including DEs and CEs, their shape is most applicable to APs, hence this modality is where the automation step is explored. Because the noise covariance, Σ, is still derived from the data, the this method is referred to as semi-supervised.

To test if this approach, which we refer to as AGMF, was suitable for extracellular AP detection, we manually identified 20 AP from the locust antenna lobe tetrode data presented in Sec. 2.3.1. The average AP SNR ranged from 1.3 to 6.1 and varied in shape. These APs were then convolved with the AGMF at 16 scales ranging from 0.5 ms to 2 ms. The resulting SNR as a function of scales ranged from 2.2 to 11.0. Overall, the SNR of the processed APs saw a median improvement of 38% ($p < 0.00004$) and the AGMF outperformed other multi-sensor AP detection approaches by yielding a significantly higher SNR when tested on the 20 APs presented here. This places the AGMF method as a good candidate for further automation and incorporation into the current detection paradigm.

Chapter 3

Classification

In most cases it is not enough to simply detect a signal of interest. Many if not most studies of biological systems strive to compare different scenarios against one another, or distinguish between the actions of several forces participating in a single phenomenon. One way to accomplish this is data classification. Just like detection, data classification in biological systems is not entirely straight forward. The signals are often noisy, and all of the parameters affecting the signal are not necessarily known. This seems like a good case for machine-learning approaches, such as evolutionary algorithms or neural-nets, however naive statistical approaches more often than not do the job with a much lower computation cost and reduced complexity. Although they are not necessarily naive, statistical approaches are the methods we concentrate on in this work.

Classification is a two part problem. First the salient features of the signal must be extracted. Then those features have to be clustered into classes. If a good features space is determined, the clusters are relatively obvious and a powerful clustering algorithm is unnecessary. On the other hand, a powerful clustering algorithm can often accurately cluster

disperse and/or overlapping features. In this work we have chosen to concentrate mostly on feature extraction, and use relatively well known algorithms for clustering.

Our first feature extraction approach was developed to classify neurons from multi-sensor extracellular neuronal recordings. A single electrode can generally record signals from 5 to 8 neurons simultaneously, although as little as 1 or as much as 20 is also possible. To really understand neural network dynamics, signals from the different neurons must be identified and delineated. The method we developed, MUSIC, focuses on finding neurons' locations from their recorded APs. Because each neuron should have a unique location, this feature extraction approach has the potential to perform better than mathematically abstract approaches. We tested MUSIC on both tetrode (4 channel) and heptode (7 channel) recorded signals. The features were then clustered using either expectation-maximization (EM) and Bayes' information criterion (BIC), or DBSCAN.

The second approach presented in this Chapter was developed to distinguish between cardiomyocyte APs, or depolarization events (DEs) affected by different drugs. We selected salient DE features – upslope, width, and downslope – based on known cardiomyocyte DE behavior, and then compared these features across different drug treatments using a Kolmogorov-Smirnov test.

This Chapter details the two methods mentioned above, and presents applications for both. Other existing techniques for each analyzed data type, as well as the motivation behind these applications are presented along with the application results for each method.

3.1 Multi-sensor Extracellular Feature Extraction: MUSIC

Depending on the electrode impedance, a single extracellular recording data set can contain signals from 1 to 20 different neurons [115]. This number can grow even larger when using electrode arrays, which often cover more surface area and have more recording sensors. The recorded APs must then be classified before yielding any useful information on neuronal function and organization. Neuronal source classification therefore represents a critical step in the analysis of electrophysiological data. However, the task itself is non-trivial. Classification is often based on spike amplitude, which is assumed to stay consistent for a single neuron over the course of recording. This assumption is often broken due to small, unintentional movements of the electrode, slight changes in the extracellular ion-concentration, shunt resistance of the electrode, and internal firing variability of the neuron. Many techniques, including the use of principle components as classification features ([25]), and EM based clustering [72], have been proposed to tackle this issue. Our approach has particularly focused on using either a combination of the EM algorithm and BIC or the density-based algorithm DBSCAN for clustering. We've also used density based algorithms such as DBSCAN [16].

Although classification methods differ in their actual clustering algorithms, the defining feature that sets most classification schemes apart is their feature selection algorithm. Choosing the correct feature largely determines the efficacy of extracellular AP classification. While commonly used abstract mathematical features, such as the principal components in PCA [32], or various template scores [42], can be used for classification purposes, their calculated features may not be unique to single neurons, and may change significantly with sensor location and across trials. Features that can reliably represent a single neuron, as well as remain invariant across trials and sensor positions are preferable for classification purposes. We believe source location is one such feature.

3.1.1 Source Location as a Feature

Neuron location is a classification feature that offers both stability and remains invariant in time and space. Source location is a superior classification feature to mathematical abstractions for several reasons. First of all, source location remains relatively constant even if the recording sensors are moved with respect to the neurons of interest. Although abrupt movements that damage or drag the tissue will undoubtedly cause neuron locations to change, slowly executed and small electrode movements should not disturb the location of the recorded neurons. Therefore neuron locations should remain constant during unexpected sensor movements which are a common occurrence during extracellular recording experiments. Secondly, source location can be used to identify and follow single neurons across trials. In chronic recordings, where neuron populations may migrate with time, this can provide both information on neural migration trends and allow for the migrating neurons to be classified accurately, and grouped consistently across many consecutive trials. Note also that localization does not require spike alignment, which decreases pre-processing time.

Due to these reasons, AP source locations were selected as our classification feature. Several neuronal source localization algorithms have been proposed and used *in vivo* [11, 14, 4, 94, 64]. These methods employed various approaches ranging from simple heuristic AP generative models [14, 4] to more biophysically realistic models, such as monopole [11], dipole [64], and line source [94] approximations. The best example is the approach taken by Chelaru and Jog, where a simplified monopole-like model is applied to the recordings and source locations are estimated by solving nonlinear systems of equations. In contrast to this numerical solution, our previous work showed that the monopole-model could be inverted exactly, resulting in a closed-form solution [52]. Unfortunately, this method is very sensitive to noise. Likewise, solutions come in pairs, one of which is spurious. Identifying which solution is the accurate one can be tedious and sometimes impossible.

Due to these limitations we explored a different localization method rooted in statistical signal processing [53]. The multiple signal classification (MUSIC) algorithm used in this work has proven effective in both electroencephalogram (EEG) and magnetoencephalogram (MEG) source localization [118], as well as in our preliminary source localization experiments with tetrodes [53]. The MUSIC algorithm is more robust against noise than the closed-form solution, and generates a single localization result, eliminating the need to identify the accurate and spurious solutions from a pair.

The work presented here concentrates on applying the MUSIC algorithm to localize multi-sensor extracellular AP for the purposes of clustering and thereby classification. The multi-sensor data explored includes a conical tetrode (4 sensors) and heptode (7 sensors).

3.1.2 Multiple Signal Classification (MUSIC)

In the most basic case, we can treat the neuron as a point source and the surrounding medium as an isotropic, homogeneous volume conductor. Although simplistic, the monopole model has often been used in application to neural source localization [14, 4, 2, 11], and will be used here as the basis of our forward model.

If signals are generated by a single source, the MUSIC algorithm models measurements from a C-sensor array, $x(t) \in \mathbb{R}^{C\times 1}$, as an output of the static linear system

$$x(t) = ms(t) + n(t) \tag{3.1}$$

where $t \in [1, T]$ is the time instant, $m \in \mathbb{R}^{C\times 1}$ is the lead field vector (LFV) [118] representing the system's response to a unitary signal input, $s(t) \in \mathbb{R}$ is the signal amplitude, and $n(t) \in \mathbb{R}^{C\times 1}$ is zero-mean noise. In the case of a single monopole-like source with current

$I(t) = s(t)$, the LFV becomes

$$m(r) = \frac{1}{4\pi\sigma}\left[\frac{1}{d_1(r)}\ \frac{1}{d_2(r)}\ \cdots\ \frac{1}{d_C(r)}\right]^T \tag{3.2}$$

where σ is the conductivity of the medium per unit length, d_c is the distance between sensor c and the source located at $r = [x,\ y,\ z]^T$, and $c \in \{1, 2, \cdots, C\}$. The components of $x(t)$ for any sensor c are then

$$x(t)_c = \frac{I(t)}{4\pi\sigma d_c(r)} + n_c(t) \tag{3.3}$$

The MUSIC algorithm proceeds by finding the source location $r^\star$ for which the LFV is most orthogonal to the noise subspace [89]. More formally, the optimal source location $r^\star$ is found by

$$r^\star = \arg\min_r \frac{m^T(r) E_N E_N^T m(r)}{m^T(r) m(r)} \tag{3.4}$$

where $E_N \in \mathbb{R}^{C\times(C-1)}$ is the noise subspace. This subspace can be obtained by the following singular value decomposition,

$$X = U\Lambda V^T \tag{3.5}$$

where $X := [x(1)\, x(2)\, \cdots\, x(t)] \in \mathbb{R}^{C\times T}$ and T is the number of samples in the time series data. If $T \geq C$ under the single-source assumption, the noise subspace can be defined as $E_N := [u(2)\, u(3)\, \cdots\, u(C)]$, where u represents the columns of U corresponding to the $C-1$ smallest singular values of X. In other words, we assume that the first singular value of X makes up the signal subspace, and the remaining values make up the noise subspace E_N. Note that σ cancels out in Eq. 3.4 so in this case localization is independent of medium conductivity.

3.1.2.1 Impedance Model

When applied to some the data in the following studies, the monopole LFV model produced very dispersed features that did not exhibit strong clustering behavior. A similar phenomenon was also encountered by Chelaru and Jog [11] in their work using a combination of source locations as classification features and a self organizing map for clustering. The underlying reason for this may be an extreme sensitivity of LFVs, especially for small d_c, which may be exacerbated in the presence of noise. The LFV was therefore regularized [18] to mitigate the noise and also incorporate the impedance of the recording channels;

$$m(r) = \lambda \frac{1}{4\pi\sigma} \left[\frac{1}{d_1(r)} \; \frac{1}{d_2(r)} \; \cdots \; \frac{1}{d_C(r)} \right]^T + (1-\lambda)[Z_1, Z_2, \cdots Z_C]^T \tag{3.6}$$

where Z_c is the impedance of sensor c, and $\lambda \in [0, 1]$ is a regularization factor. The value of λ that produced optimally clustered features was empirically derived as $\lambda = 0.5$. The impedance values used were measured for each sensor prior to recording in 9% saline at 1 KHz.

In this case σ does not cancel out so localization is in small part dependent on the medium conductivity. We have assumed a standard conductivity for cerebral spinal fluid, $\sigma = 0.3 \times 10^{-6}$ S/μm [36]. Perturbation of this parameter slightly affects the localization results, but does not impact the feature clustering.

Although this new model is likely inaccurate for the purposes of real source localization, the regularization forces very tight clusters in the feature space, enhancing clustering performance. That being said, each cluster tends to be closely associated with a single sensor, which means that the maximum number of clusters the algorithm can identify is C, the number of sensors. Although we did not run into this issue, as the largest number of units identified in a given recording was 6 while using a 7 sensor electrode, it is a possible limitation of the approach.

3.1.3 Proof of Concept

Originally the MUSIC localization approach was intended both as a classification feature, as well as a tool to determine true neuron locations from multi-sensor extracellular recordings. To determine if the algorithm would be capable of localizing an electrical source we first tested it in a contrived setting, using a stimulator placed in artificial cerebrospinal fluid (aCSF) and four recording micropipettes. The location of the source was estimated with an accuracy and precision of $\sim$4 μm and $\sim$7 μm, respectively. These results suggested that *in vivo* resolution of individual neuronal sources was feasible using MUSIC as a feature. Although ultimately the MUSIC algorithm proved to be unsuccessful in true neuronal source localization, these results led to our further studies on using the MUSIC algorithm to extract classification features that did not necessarily correspond with physical neuronal locations.

The following work presents experimental verification of the MUSIC algorithm's ability to localize electrical sources in aCSF from tetrode recorded signals.

3.1.3.1 Methods

Our experimental setup utilized four glass micropipette electrodes for recording and a single micropipette electrode for stimulation. MUSIC estimated source locations were compared to the true source location, as determined from a microscope image. Since the precision of the estimated source locations ($\sim$7 μm) is smaller than the diameter of a typical soma, we hypothesized that this method can resolve individual neuronal sources based on their recorded APs.

Data Collection Five glass micropipettes (Fig. 3.1) filled with aCSF were placed in a recording chamber maintained at 32°C. Micropipettes 1-4 served for recording, while micropipette S served as a stimulator. The recording chamber was immersed in aCSF consisting

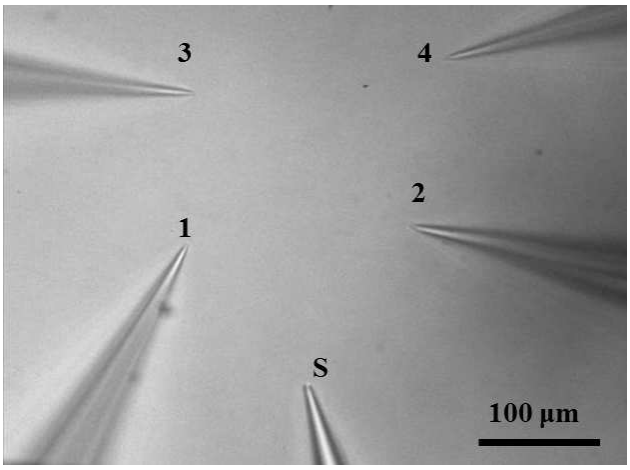

Figure 3.1: Microscope image of the experimental preparation. The micropipette tips lie in the microscope focal plane. The micropipettes were subsequently brought closer by a micromanipulator (see Fig. 3.2).

of 127 mM NaCl, 26 mM NaHCO3, 3.3 mM KCl, 1.24 mM KH2PO4, 1.0 mM MgSO4, 1.0 mM CaCl2, and 10 mM glucose. Electric field potentials recorded by the four micropipettes were acquired through an Axopatch 700B amplifier (Molecular Devices, Union City, CA) with a sampling rate of 20 kHz. The stimulating micropipette generated a train of ten 38-μA pulses, each pulse lasting 10 ms, followed by a 40-ms pause. To reduce the effects of noise, a total of seven such trains were recorded. Images were collected during data acquisition with an up-right microscope (Nikon, Tokyo, Japan) and captured with a cooled CCD camera (iXon DV885, Andor Technology, Belfast, UK). The experimental procedure was performed in accordance with the University of Tokyo guidelines.

MUSIC Application and Performance Analysis The recorded pulse trains were aligned to the onset of the first pulse, corrected for dc offset, and averaged over the seven epochs. Source location was then estimated using the MUSIC algorithm, and the accuracy of the estimated source location was quantified by calculating the estimation error (bias) defined as:

$$\varepsilon \triangleq \|E\{r^*\} - r_t\| \tag{3.7}$$

where $\|\cdot\|$ represents the Euclidean norm, E is the expectation operator (average), $r^* \in \mathbb{R}^{3\times 1}$ are estimated source locations, and $r_t \in \mathbb{R}^{3\times 1}$ is the true source location. Similarly, the precision of the estimate is quantified by a standard radius (generalized standard deviation) defined as:

$$\delta \triangleq \sqrt{\delta_x^2 + \delta_y^2 + \delta_z^2} \tag{3.8}$$

where δ_x, δ_y and δ_z are the standard deviations of the x, y and z components of r^*, respectively.

3.1.3.2 Results

The recorded pulse trains were processed and analyzed as described in Sec. 3.1.3.1. As expected, the strongest signals were recorded by sensors 1 and 2, which lied closest to the source (Fig. 3.2 (Left)).

The 10-ms data segments, corresponding to each of the 10 pulses were used as input to the MUSIC algorithm, yielding one estimated source location for each pulse (Fig. 3.2 (Center)). The MUSIC-derived source location average over the 10 pulses was (-6.45, 43.02, 4.70) μm (B in Fig. 3.2 (Left)), with the origin defined at the source, S. This corresponds to an error, ε, of 43.75 μm. While Fig. 3.2 (Center) shows that the solutions were consistent in the x-y plane ($\delta_x = 0.03$ μm, $\delta_y = 0.15$ μm), the variance along the z-axis was high ($\delta_z = 22.25$, μm), yielding an overall standard radius, δ, of 22.25 μm.

This relatively large bias ($\sim$44 μm) is consistent with our previous study (bias: $\sim$41 μm) [53]. It indicates that the modeling assumptions and, in turn, the constraint imposed on the LFV (3.2) may not hold. In the simplest scenario, this bias can be attributed to medium inhomogeneity and mitigated by allowing each source-sensor path to have a different con-

ductivity value. This redefines the LFV (3.2) as:

$$m(r) = \frac{1}{4\pi\sigma}\left[\frac{1}{k_1 d_1(r)}\ \frac{1}{k_2 d_2(r)}\ \frac{1}{k_3 d_3(r)}\ \frac{1}{k_4 d_4(r)}\right]^{\mathrm{T}} \tag{3.9}$$

where k_i (i = 1, 2, 3, 4) is a constant making each conductivity value a multiple of some baseline conductivity σ. This multiplier, referred to as an inhomogeneity correction factor (ICF), can be found experimentally, provided the distance, d_i (i = 1, 2, 3, 4), is known. By assuming that one of the stimulator-sensor paths has the baseline conductivity, $k_4 = 1$, the remaining ICFs can be found using (3.1) and (3.9), and taking the expectation of x over the noise distribution:

$$k_c(t) = \frac{d_4 E\{x_4(t)\}}{d_i E\{x_i(t)\}}, \quad c = 1,\ 2,\ 3 \tag{3.10}$$

where $x_c \in \mathbb{R}$ is the cth component of x (for a given channel c). Although k_c depends on time, we have shown that its values remain fairly stable over time [53] and can be estimated by taking the median value of $k_c(t)$. This yielded $k_1 = 1.20$, $k_2 = 1.04$ and $k_3 = 0.99$. The MUSIC algorithm can then be executed using the new LFVs (Eqs. 3.9- 3.10).

The localization results significantly improved upon ICF correction. The average estimated source location was (0.20, 0.61, 3.78) μm (A in Fig. 3.2 (Left)), with corresponding error $\varepsilon = 3.83$ μm. This is also highly consistent with our previously reported results (error: $\sim$3 μm) [53]. Likewise, ICF adjustment substantially reduced localization variance ($\delta_x = 0.40$ μm, $\delta_y = 3.50$ μm, and $\delta_z = 6.27$ μm), yielding a standard radius, δ, of 7.20 μm. The clustering of these sources is shown in Fig. 3.2 (Right).

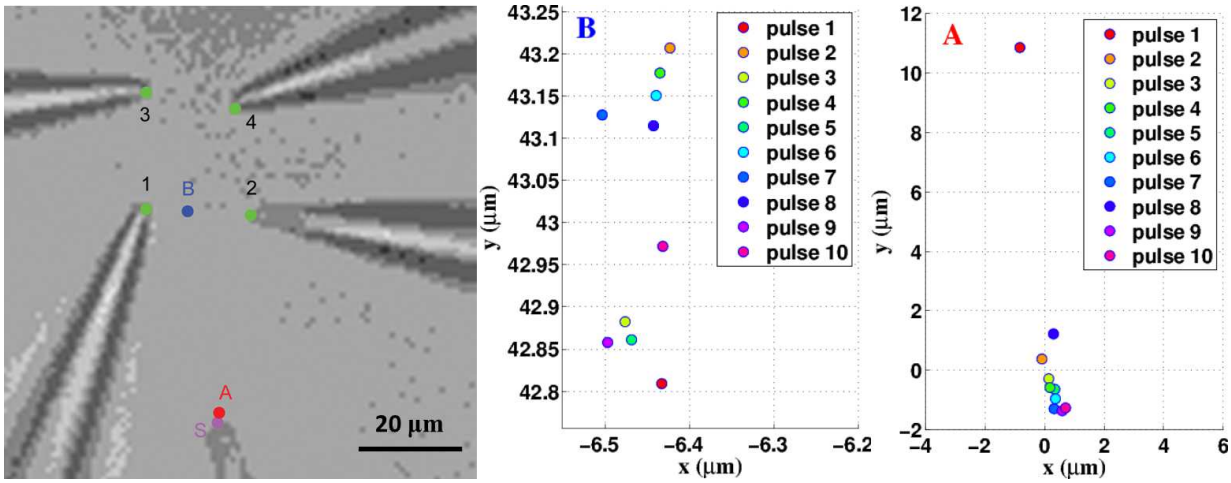

Figure 3.2: (Left) The final sensor-stimulator configuration. The sensors (the tips of micropipettes 1-4) are marked in green, and the source (the tip of the stimulating micropipette) is marked in magenta. The average of estimated sources is marked by B. The equivalent point after ICF adjustment is marked by A. (Center) The estimated locations of 10 sources corresponding to the 10 pulses before ICF adjustment. (Right) The equivalent plot for the pulse train after ICF adjustment.

3.1.3.3 Discussion and Conclusion

The sole purpose of ICF adjustment in our analysis was to validate the performance of the localization algorithm. Since determining ICFs relies on known distances between the sensors and the source, ICF adjustment cannot be performed on blind data collected *in vitro* or *in vivo*. Source localization based on unadjusted signals, however, is likely to result in a relatively large bias and increased variance.

As suggested in Section 3.1.3.2, the bias is believed to have been primarily caused by medium inhomogeneity, although other factors such as medium anisotropy or non-uniform sensor impedances could play a role. However, preliminary evidence suggests that impedance inconsistency is not likely to be the cause, as similar bias values have been observed in our previous study [53], performed using a single sensor placed at four recording locations.

Another potential source of bias could be a failure of the monopole model to accurately describe the source. Our preliminary *in silico* work [53], as well as *in vivo* studies performed

by Mechler and Victor [63], indicate that discrepancies between monopole-generated signals and measured neuronal signals often result in localization solutions biased toward the sensors. Furthermore, the main cause for the relatively low precision of source localization prior to ICF adjustments was caused mostly by obtaining source locations that were geometrically mirrored on either side of the sensor plane, with the z-coordinates across 10 localization trials clustered at either $\sim$23 μm or -23 μm. Other work performed using this algorithm (not discussed in this dissertation) also lacked effective z-axis resolution even with a 3D sensor arrangements.

This leads us to conclude that the MUSIC algorithm presented here is too prone to bias for accurate neuronal source localization. However, the localization precision, especially in the x-y plane shows the MUSIC algorithm's ability to distinguish electrical sources, using 4 sensor recorded signals. Although a bias may appear, the precision of localization indicates that this approach may be able to distinguish independent neurons from multi-sensor multi-unit extracellular recordings, and makes it a good candidate for use in spike sorting.

Our further work therefore focuses on testing MUSIC as a feature extraction method for the purposes of multi-sensor spike sorting.

3.1.4 Applications

The MUSIC algorithm for feature extraction was applied to two different data sets - one collected with a tetrode, and another collected with a heptode. The first application, using data from the tetrode, applies the base monopole model (Sec. 3.1.2) and clusters the resulting features with an EM-BIC based approach. Although no ground-truth was available, the resulting classified waveforms were compared, and inter- and well as within-class analyses were preformed to test classification efficacy. The second application, using data from the heptode, uses the impedance model (Sec. 3.1.2.1) and then clusters the resulting features

with DBSCAN. In this case the waveforms were qualitatively inspected for classification efficacy across several different experimental paradigms, and t-tests were used to compare waveforms across different classes.

3.1.4.1 Monopole Model with Tetrode

Extracellular APs must be classified before they can yield any useful information on neuronal function and organization. Neuronal source classification therefore represents a critical step in the analysis of electrophysiological data. This study demonstrates the efficacy of a multi-sensor AP classification scheme using source location as a classification feature.

The data and results presented here are adapted from Szymanska et al. 2013 [99]. Briefly, localization was performed using the multiple signal classification (MUSIC) algorithm. Six distinct source neurons were classified from 20 seconds of extracellular tetrode recordings. On average, 89.5% of the waveforms making up each class matched the shape of the average class waveform. These results indicate that this classification scheme can successfully identify individual neurons from multi-sensor AP recordings.

Methods

Data Collection The data used to investigate this approach is the same data as in Sec. 2.3.1, detailing the matched-filter approach for extracellular action potential detection. The data is available online [78], and additional details about the data collection procedure beyond that provided in Sec. 2.3.1 are described in Pouzat et al. 2002 [79]. Briefly, a planar silicon probe was placed below the surface ($\sim$50-100 μm) of an adult locust's antennal lobe and used for recording. Recordings were sampled at 15 KHz and bandpass filtered from

300 Hz to 5,000 Hz. A total of 20 seconds of data from four of the probe tip sensors was provided.

Data Analysis Spike detection was performed using the supervised matched filter for multi-sensor data presented in Sec. 2.3.1. Half of the collected data was used for training, and the remaining half was used for further analysis. 25 APs, 2.7 ms each, were selected from the training data, and used to generate the matched template, s, and 20 noise samples, roughly 50 ms each, were used to generate the noise covariance matrix, Σ. Like in Sec. 2.3.1, the noise statistics were assume to be spatially white to simplify the noise covariance calculation. The matched filter output was thresholded at three standard deviations above the noise mean, $a = 3$.

The source of each detected AP was then localized using the MUSIC algorithm, with a monopole source model. MUSIC-derived locations from each detected AP were then classified using an EM algorithm. The EM algorithm assumes Gaussian distributed clusters for APs coming from a specific neuron, and a uniformly distributed cluster of outliers. The EM algorithm was used to group points into several different cluster models, and the optimal cluster model, or number of clusters, was determined by maximizing the BIC across all models [90]. For a detailed derivation of this classifier please refer Appendix B.1.

After classification each cluster was analyzed for average location and spread. The spread was quantified by a standard radius, which is the norm of each cluster's x, y, and z standard deviations. The APs from each cluster were aligned to their peak values and averaged to demonstrate the representative waveforms for each cluster. Waveform signatures, representing the relative signal power across the four sensors, were then calculated for each AP, and used to determine the within-class consistency between the waveforms.

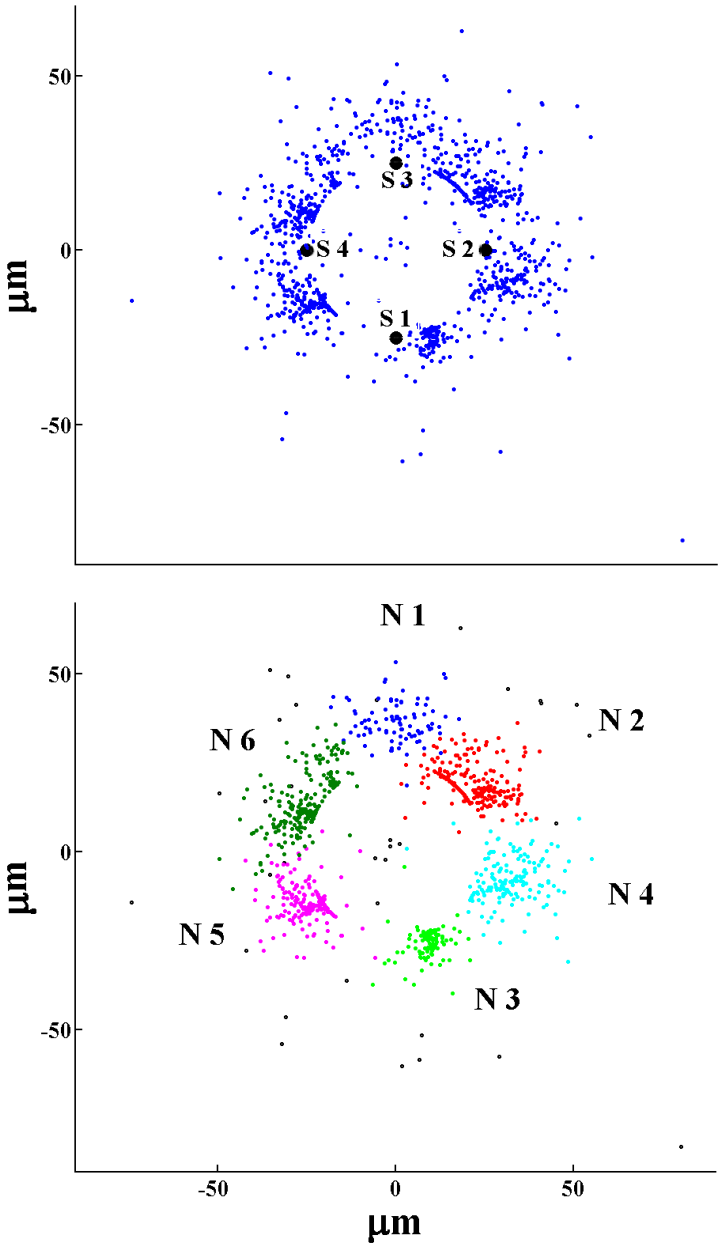

Figure 3.3: (*Top*) MUSIC source localization results for all 1040 detected APs (blue). Sensor locations are depicted in black and marked S1-4. (*Bottom*) Classification results for AP source locations shown above. Black points represent outliers, and all other colors represent distinct source location clusters, labeled N1-6.

Table 3.1: Location, standard radius, and classification accuracy of each class N1-6

Source		N1	N2	N3	N4	N5	N6
Location	x (μm)	0.4	21.3	9.5	32.1	-24.8	-25.4
	y (μm)	36.4	19.3	-25.7	-8.0	-14.4	13.5
	z (μm)	0.5	-0.1	-0.1	-0.7	0.2	-0.2
Std. Radius (μm)		11.5	10.1	6.7	12.3	9.0	11.2
Accuracy (%)		98.1	77.5	95.4	98.1	90.1	77.2

Results All 1040 APs detected in the 20 second data stream were successfully localized using MUSIC. The AP locations, projected onto the x-y plane, are shown in Fig. 3.3(*Top*). This data was classified yielding 6 distinct location clusters, centered at (0.4 36.4 0.5), (21.3 19.3 -0.1), (9.5 -25.7 -0.1), (32.1 -8.0 -0.7), (-24.8 -14.4 0.2), (-25.4 13.5 -0.2) μm with standard radii of 11.5, 10.1, 6.7, 12.3, 9.0, and 11.2 μm respectively (Fig. 3.3 (*Bottom*), Table 3.1). Only 34 source locations were classified as outliers, representing 3% of the data set.

The underlying AP waveforms representing each source location in a given cluster were analyzed to assess classification efficacy. The average waveforms are unique to each location and distinguishable from each other (Fig. 3.4). This is strong evidence that each cluster represents a unique neuron.

Although the average waveforms are unique, all six have relatively high standard deviations at the AP peaks (up to 4.9 standard deviations of the noise). This is not surprising as the waveforms were crudely aligned by peak AP values, greatly increasing the peak variance. Additionally, biological noise is known to increase during spikes due to the correlated activity of nearby neurons [30]. To determine more conclusively if this high waveform variance implies misclassification, waveform signatures, defined here as the ranking of signal power across the four sensors for each AP, were compared within clusters.

For clusters N1, N3, N4, and N5, the waveform signatures are consistent with the average waveform signature among 98.9%, 95.4%, 98.1%, and 90.1% of the APs in each cluster, respectively (Table 3.1). This implies that the variance in AP waveforms across these clusters is most likely due to mis-alignment or biological noise. It is reasonable to conclude that these clusters represent unique and singular neurons.

The remaining two clusters were slightly less consistent, with 77.5% and 77.2% of the APs in N2 and N6, respectively, matching their average waveform signatures. For N6 the remaining 22.8% of waveforms match the waveform signature of N1, and therefore seem to be misclassified. N2 on the other hand had a broad spectrum of APs not matching the average waveform. This may imply that some APs classified as N2 account for a different nearby neuron, or a superposition from two or more neurons. Although these clusters were less internally consistent, the results still indicate that they represent distinct neurons.

Overall, classification based on our feature extraction method successfully differentiated 6 unique and distinct neurons from tetrode recorded APs.

Discussion and Conclusion The misclassification observed in clusters N2 and N6 may be due in part to the use of only 4 recording sensors, the minimum necessary for MUSIC localization. The use of arrays with more than 4 sensors may decrease localization error and misclassification rates. As 4 sensor localization is sensitive to outlying data, a larger number of sensors may mitigate the effects of noise and decrease localization variance.

A notable characteristic of this feature extraction scheme is its limited sensitivity to noisy or outlying signals. Unlike the approach presented by Chelaru and Jog [11], where 39% of recorded spikes were filtered out and discarded as outliers prior to analysis, the feature extraction scheme presented here did not filter out any spikes prior to analysis. Furthermore, only 3% of the classified spikes were identified as outliers, conserving most of the

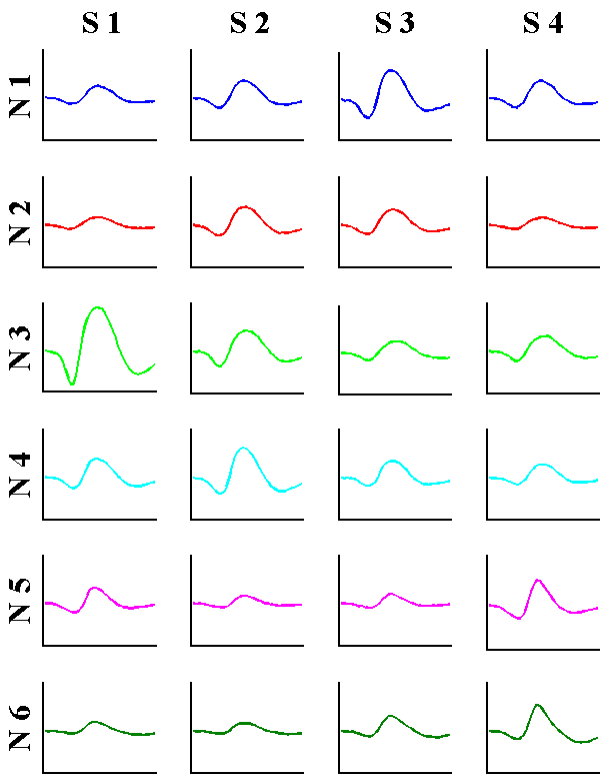

Figure 3.4: Average waveforms for each cluster shown in Fig. 3.3 (*Bottom*). Waveforms are color matched to the clusters. Note that each cluster has a unique and distinguishable waveform signature.

detected APs. This can increase the reliability of further analysis, and improve the amount of information that can be gathered from a given recording session.

Overall, 6 distinct AP sources were successfully classified, with an average accuracy of 89.5%, using MUSIC-derived source location as a classification feature. Furthermore, the classification scheme used here does not rely on any prior knowledge of AP class characteristics, and does not require the signals to be aligned prior to classification. This decreases the amount of signal processing necessary for classification, allows a broader range of units to be identified, and permits the algorithm to be completely unsupervised. Given our results, this technique presents itself as a strong candidate for broad use in extracellular signal analysis.

3.1.4.2 Impedance Model with Heptode

As mentioned previously spike sorting or neuronal source classification is crucial to the analysis of electrophysiological data. This study demonstrates the efficacy of a multi-sensor AP classification scheme using a proxy for source "location" as a classification feature (Sec. 3.1.2.1). Localization was performed using the multiple signal classification (MUSIC) algorithm, with the impedance model (Sec. 3.1.2.1). The approach was tested on extracellular heptode recordings, which we believe will provide better resolution and stability against noise than the previously tested tetrode recordings. Five to six statistically distinct source neurons could be classified from a given heptode recording (Appendix A.3), across the 3 independent recordings tested. This shows that the method is effective for multi-sensor spike sorting.

The results presented here were used in Szymanska et al. 2017 [101], but were not presented in detail.

Methods The data collection procedure for this work is detailed in Appendix A.3 as part of the study described fully in Chapter 4. Here we will summarize the data and methods pertinent to spike sorting.

Two male Wistar rats participated in the study. Extracellular heptode recordings were made through a 1-2 mm burr hole above M1 of the right hemisphere. A screw electrode placed over V1 of the right hemisphere served as the ground. A motorized headstage was used to lower the heptode into the cortex until a high activity and signal-to-noise ratio recording site was reached ($\sim 700 - 900\mu$m into the cortex). Once the heptode was in position 1-2 min of spontaneous neuronal activity were recorded at three monotonically increasing isoflurane anesthesia levels. For the purposes of this work, we will concentrate on the lowest level of anesthesia, where the bulk of spiking activity occurred.

A total of three experiments were performed. The first two experiments were performed on a single animal at two different recording sites, one 170μm below the other. These locations are presumed to be far enough apart to be recording from a new population of neurons. These two experiments will be referred to as experiment 1.1 and 1.2, and were performed at 2.0% and 1.5% isoflurane, respectively. The third experiment was performed on the second animal at 1.5% isoflurane and will be referred to as experiment 2.1. For more details please refer to Appendix A.3.

Individual APs were detected from the heptode data using the generalized matched filter described in Sec. 2.3.1 and Szymanska et al. (2014). Manually selected APs and noise segments from each data set served as the training data, and the filter thresholds used ranged from 4.5-6.5 median standard deviations above the median filter output [102], depending on the data SNR. The APs were then classified using the MUSIC algorithm impedance model for feature extraction (Sec. 3.1.2.1), and DBSCAN for clustering [16] (Appendix B.2).

The classified APs were then aligned to their peak values and averaged to demonstrate the waveform signature for each cluster. The waveform signatures were qualitatively compared to assess classification efficacy. Peak AP values were compared using t-tests (95% confidence level) across each recording sensor to quantitatively assess if classes were statistically distinct.

Results Almost all of the detected APs were successfully localized using the MUSIC impedance model for experiment 1.2; only 84 out of 1249 detected APs (7%) could not be localized. For experiment 1.1 only half of the detected APs (596 out of 1184) were successfully localized. Inspection of the un-localized waveforms revealed that almost all consisted of overlapping APs. Since the overlapping APs likely originated from different neurons, it is not surprising that MUSIC localization could not converge on a solution. Similarly only 2,437 of the 5,647 detected APs (43%) in experiment 2.1 were successfully localized. Here about half of the un-localized waveforms contained overlapping APs, where as the other half

did not contain any obvious imperfections aside from a lower average SNR than the localized APs. The localization results are shown in Fig. 3.5 **A** (*Left*).

As shown by the localization results in Fig. 3.5 **A** (*Left*), the MUSIC impedance model has "localized" the APs to be tightly clustered around the heptode sensors. This is in stark contrast to the more diffuse localization results presented in Sec. 3.1.4.1, where only the base monopole model was used. As mentioned in Sec. 3.1.2.1, this an expected result as the assumption that each sensor is connected in series with the neuron is likely inaccurate for the purposes of real source localization, but does constrain the AP sources to very tight cluster in the feature space. The clusters therefore become extremely obvious.

Fig. 3.5 **A** (*Right*) shows the DBSCAN clustering results for the location features on the left. As can be seen in the figure, DBSCAN does a good job clustering the features. DBSCAN identified 35 outliers out of the 596 features (6%) for experiment 1.1, 42 outliers out of the 1123 features (4%) for experiment 1.2, and 162 out of the 2275 features (7%) for experiment 2.1. Some of the outliers could be classified with slightly more lax parameters for the density based algorithm, however, here we selected to be very conservative, ensuring that only the best APs were classified.

Five, six, and six classes were identified for experiments 1.1, 1.2, and 2.1 respectively. Fig. 3.5 **B** shows the waveform signatures for each class identified in the three experiments. The waveform signatures are clearly distinct between the different classes. Furthermore, as the plotted 95% confidence intervals show (they are almost invisible in the Figure), the APs in each class were quite consistent. T-tests (95% confidence interval) were performed on the peak values of the waveforms for each channel to quantitatively determine if the classes within a given experiment were statistically distinguishable in the time domain. Each channel was compared to the corresponding channel for another class. For example when comparing C1 and C2 for experiment 1.1, 7 t-tests would be performed comparing the peak AP values between C1 CH1 / C2 CH1, C1 CH2 / C2 CH2 and so on. In all case, at least 5 of the

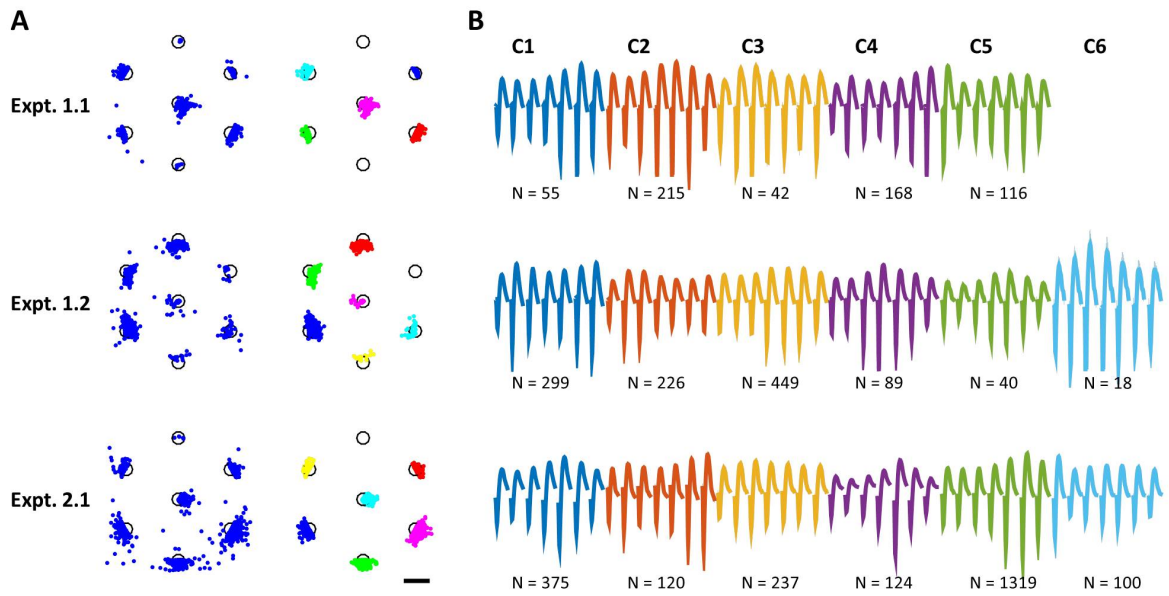

Figure 3.5: Spike sorting results using the MUSIC impedance model and DBSCAN. **A** An aerial view of the feature space and clustering results. The black bar in the bottom right corner represents 10 μm. (*Left*) Results of MUSIC localization for each experiment. The black circles represent the heptode sensors, and the blue dots are MUSIC determined source locations for each localized AP. (*Right*) DBSCAN clustering results for the features presented on the left, with outliers removed. **B** Corresponding average waveforms for each cluster identified in **A**. The number of APs in each class is listed below the waveform. Although 95% confidence intervals are plotted as shaded regions around the waveforms, they are so small that they can not be seen for most of the waveforms. T-tests revealed that the waveforms within a given experiment are statistically distinct across classes.

channels showed a statistical difference between the classes, and in most cases all 7 were statistically different. The rare channels that were not statistically different were always the channels with the lowest signal amplitude. To our mind, this indicates that all of the classes in a given experiment are statistically distinguishable from each other.

Discussion and Conclusion The MUSIC impedance model algorithm successfully classified 5-6 source neurons for the three experiments presented here. Although not all of the sources could be localized using the MUSIC impedance model, most of the feature extraction outliers were due to superimposed APs, which were unfortunately very common in these data sets. Other APs that could not be localized suffered from low SNR. These are both

clear limitations of the technique, especially compared to the base MUSIC monopole model, which was able to localize all of the APs in the tetrode data set presented in Sec. 3.1.4.1. This discrepancy could be a function of the data, since the data sets tested here contained many superimposed APs and the one data set presented in Sec. 3.1.4.1 did not. Either way it appears that this model is more sensitive to low SNR. On the other hand, the features extracted using the MUSIC impedance model are much more tightly clustered than those achieved using the monopole model. As can be seen by the very small 95% confidence intervals plotted in Fig. 3.5, the classes are very consistent, and appear to be less prone to misclassification.

Another possible limitation of this approach, also mentioned in Sec. 3.1.2.1, is the natural upper-bound on the number of classes that can be identified. Because the source locations tend to cluster around a specific sensor, the maximum number of clusters that can be identified depends on the number of recording sensors. In this case, this limitation proved not to be an issue, as the maximum number of identified classes was 6, and a 7 sensor array was used. However, it is a potential shortcoming of the technique compared with the MUSIC monopole model which is not limited by this phenomenon.

It is worth noting that the data presented here was also treated with the MUSIC monopole model (data not shown). Although the detected APs could be localized for the most part, the features space was very crowded and the features were diffuse, therefore no individual clusters could be identified. The data was also treated with several common spike sorting techniques including WaveClus [80], and KlustaKwik [32], as well as a newer method the Matched Subspace Detector (MSD) [117]. All were used with default setting. The results for all three approaches was one large cluster containing roughly 50% of the APs. This cluster contained several AP waveform shapes and was clearly riddled with misclassification. The remaining APs were for the most part identified as outliers, with an occasional small ($< 100APs$) second class. There approaches were therefore not viable for spike sorting

of heptode data. We therefore conclude that although the impedance model suffers form several limitations, it appears to be more apt than both the monopole model as well as existing techniques at dealing with certain data sets.

Overall, the MUSIC impedance model was effective at classifying APs from extracellular heptode data. Furthermore, both qualitative analysis as well as t-test showed that the AP waveforms were consistent within each class, and that the waveforms were statistically different between classes.

3.1.5 Limitations

As seen in Sec. 3.1.4, MUSIC based feature extraction suffers from several limitations. One of the major limitations is the inability of the algorithm to always come up with a feature. The number of these kinds of outliers ranges from 0% to 50%. Although, this is not ideal, it is consistent with other existing methods, like KlustaKwik and WaveClus which routinely reject about 50% of the data as outliers. Furthermore, unlike these existing techniques our MUSIC based approach is the only one specifically tailored for multi-sensor data, as opposed to single sensor data. Although both KlustaKwik and WaveClus can be adapted to multi-sensor data, it is our experience that the performance degrades and becomes less reliable. For example, given the heptode data presented in Sec. 3.1.4.2, both KlustaKwik and WaveClus were unable to identify more than 2 distinct classes from any of the recordings and removed up to 80% of the data as outliers. Our design therefore makes this approach uniquely suited for multi-sensor data containing many units with similar waveform signatures that are likely to be misclassified or rejected as outliers by other techniques.

Another limitation of this approach is that the z-axis, relative to the electrode, is notoriously incorrectly estimated. All of the sources in the case of the monopole model are contained right around $z = 0$, and all of the sources in the case of the impedance model are confined to

the z-coordinates of the recording sensors. This indicates the MUSIC derived locations are only related to the real source locations, and do not reflect the true neuronal soma positions. Nevertheless, the exact locations of the neurons are not necessary for spike sorting. As long as the relative locations in the x-y plane allow for accurate clustering, which in the cases presented here seems to be true, sources can be accurately sorted even if the MUSIC features are a distorted representation of the true neuronal location space.

Lastly, although the impedance model creates very tight clusters, even when the base monopole model cannot (and sources cannot subsequently be sorted), it suffers from a major limitations: the number of sources cannot exceed the number of recording channels. This was discussed in detail in Sections 3.1.2.1 and 3.1.4.2. Although this limitation proved not to be a major issue in our studies, it may be more problematic if the recording electrode is a tetrode containing only 4 channels. The impedance approach is therefore best if used with a high number of recording sensors.

3.2 Other Statistical Methods: DEC

Although many sophisticated statistical methods exist for distinguishing between various data sets and variables, their complexity must be weighed against added value and improved efficacy. In laymen's terms, the simples approach is often the best. The work presented here uses this philosophy to distinguish cardiomyocyte DEs being affected by two different drug applications from controls.

The work presented here is adapted from Szymanska et al 2016 [103] and the data was also presented in Sec. 2.3.2 detailing the MaD detector for voltage sensitive dye (VSD) imaged cardiomyocyte data. Briefly, this study used 2-photon microscopy of fluorescent VSDs to capture the membrane voltage of actively beating human induced pluripotent stem cell-

derived cardiomyocytes (hiPS-CMs). We built a custom and freely available Matlab software to quantify, and compare DEs of hiPS-CMs treated with the β-adrenergic drugs, propranolol and isoproterenol. The classification approach aims to distinguish the drug treated cardiomyocyte APs by extracting salient AP features, such as upslope, width, and downslope, and then comparing them across drug treatments using a Kolmogorov-Smirnov (K-S) test. The software, depolarization event comparison or DEC, was able to distinguish control DEs from drug-treated DEs both immediately as well as 10 min after drug administration.

3.2.1 Background on Cardiomyocyte Electrophysiology Analysis

As mentioned in Sec. 2.3.2, the popularity of human cardiac cell research as well as the expanded availability of *in vitro* cardiomyocytes given various stem cell differentiation techniques, has increased the need for new methods that can assess the electrophysiological effects of various drug compounds on cardiac cells *in vitro*. Image-based tools for assessing cardiomyocytes, such as voltage-sensitive dye (VSD) imaging, are proving to be particularly popular [34, 120, 24, 33, 55] for these types of studies.

To quantitatively assess drug induced electrophysiological effects from VSD imaged cardiomyocytes, we built a custom Matlab software for depolarization event comparison (DEC). Most other processing tools for this type of data concentrate on signal detection [97, 50, 8, 3, 112], and do not provide further analysis tools for quantitatively assessing changes in tissue or cell activity with varying experimental parameters (such as drug administration). Our previous work has analyzed hiPS-CM electrophysiology by performing supervised machine learning on pre-defined DE parameters [34]. Here detected DEs are compared using a K-S test across treatments and time points. The chronotropic drugs propranolol and isoproterenol were selected to validate this VSD-based approach and the corresponding analysis software. This method allows for quantitative assessment of the heterogeneity of DEs at pre-

cise locations on the membranes of actively beating cardiomyocytes, as well as quantitative assessment of how a given drug affects the DE shape.

3.2.2 Methods

3.2.2.1 Data Collection

The data used here is the same as that presented in Sec. 2.3.2. The full data collection procedure is described in Appendix A.1 as well as Szymanska et al. 2016 [103] and Heylman et al. 2015 [34]. Briefly, spontaneously beating hiPS-CMs were stained with VSD, confirmed to still be spontaneously beating, and treated with the β-adrenergic drugs propranolol and isoproterenol. The cultures were imaged immediately after addition of drugs (less than 60 sec of exposure) and again 10 to 15 min after addition to ensure complete exposure. One culture was treated with propranolol, and two cultures were treated with isoproterenol. Two cultures were left untreated and imaged as controls. This accounts for a total of 7 data traces, from 5 cell cultures. The imaging data was then processed to extract fluorescence intensity traces along a given cell membrane, and filtered to remove photobleaching artifacts.

3.2.2.2 DE Comparison (DEC) Analysis

Once the DEs were detected from each data trace, using the technique presented in Sec. 2.3.2, the individual spikes were extracted from the data and normalized such that the average DE for any given data trace had a minimum value of 0 and a maximum value of 1. This approach allowed us to preserve within data trace variations around the average DE, while also normalizing DE amplitudes across different data traces. We chose to normalize the DEs because their amplitudes are highly dependent on photobleaching, exact position in the imaging plane, as well as how much of the VSD each membrane initially absorbed. It is

therefore unreliable to compare DE amplitudes across treatments and time. Each DE was defined by a window consisting of 800 time points ($\simeq$ 0.48 s) centered on the maximum value of the DE. The average waveforms were calculated for each data trace. The full-width half-max (width), the positive slope at half-max (upslope), and the negative slope at half-max (downslope) were calculated for each identified DE. These parameters were compared as a function of drug treatment and time elapsed after drug treatment using a two-sample Kolmogorov-Smirnov (K-S) test at a 5% significance level.

3.2.3 Results

As mentioned in Sec. 2.3.2, A total of 7 data traces, from 5 cell cultures, were tested. The first 5 conditions are a control (cell culture 1), immediately after addition of isoproterenol (cell culture 2), 10 min after addition of isoproterenol (cell culture 2), immediately after addition of propranolol (cell culture 3), and 10 min after addition of propranolol (cell culture 3). These conditions will now be referred to as Control-1, Iso-1 0min, Iso-1 10min, Pro-1 0min, and Pro-1 10min, respectively. In order to better test DEC in low-SNR environments we also provide results for a low-SNR control data case (cell culture 4), as well as a low-SNR 15 min after addition of isoproterenol case (cell culture 5). These two conditions will now be referred to as Control-2 and Iso-2 15min, respectively.

The DEs detected at optimal thresholds as determined in Sec. 2.3.2 were used for spike analysis. All p-values associated with this analysis, as well as the average DEs of the treatments being compared, are shown in Figs. 3.6 and 3.7. Two DE groups were considered distinguishable if at least one of the metrics being measured (DE widths, upslopes and downslopes) were statistically different between the two groups, as determined by KS-tests. The drug treatments were first compared with controls, and then across time elapsed since drug administration.

The higher SNR drug treated data (Iso-1 and Pro-1) were compared to the higher SNR control (Control-1). The Iso-1 0min drug treatment was statistically distinguishable from Control-1 in width (p-value = 3.6×10^{-5}), upslope (p-value = 7.9×10^{-7}), and downslope (p-value = 6.4×10^{-6}). This difference was maintained 10 min after isoproterenol was administered (Iso-1 10min) (width p-value = 5.0×10^{-8}, upslope p-value = 0.02, downslope p-value = 1.6×10^{-7}). DEs measured immediately after propranolol administration (Pro-1 0min) were also already statistically distinguishable from Control-1 in width (p-value = 7.0×10^{-11}), upslope (p-value = 2.5×10^{-3}), and downslope (p-value = 1.1×10^{-4}). This difference was also maintained 10 min after propranolol was administered (Pro-1 10min) (width p-value = 1.5×10^{-7}, upslope p-value = 1.1×10^{-8}, downslope p-value = 5.5×10^{-7}). These results (shown in Fig. 3.6) indicate that this analysis method can distinguish DEs from control and drug treatments even less than a minute after the drug is administered (0min cases). The distinctions are also maintained 10 min after drug administration.

To show that the method is also applicable in a low-SNR setting, we performed a DE comparison analysis on Control-2 (average SNR = 3.07) and Iso-2 15min (average SNR = 0.63). The Iso-2 15min drug treatment was statistically distinguishable from Control-2 in width (p-value = 0.02), but not in either upslope (p-value = 0.68) or downslope (p-value = 0.25). These results are shown in Fig. 3.7. We would expect the DE width difference to be most pronounced among these two cases. This is reflected in the results. As SNR is decreased the DE widths remain the only distinguishable parameters, while upslope and downslope no longer differ between the two cases.

The drug treatments were then compared in time to evaluate if the 0 min cases, exhibiting initial shock from the drugs, could be distinguished from the 10 min cases, which should exhibit a more stabilized response. The Iso-1 0min case was statistically distinguishable from the Iso-1 10min case in all three parameters (width p-value = 2.2×10^{-4}, upslope p-value = 1.6×10^{-5}, downslope p-value = 8.3×10^{-4}). The Pro-1 0min case was statistically

distinguishable from the Pro-1 10min case in width (p-value = 0.01) and upslope (p-value = 1.4×10^{-3}), but was not statistically different in downslope (p-value = 0.28).

In summary, KS-tests of DE widths, upslopes and downslopes revealed that DEs immediately after isoproterenol administration are distinguishable from controls. Similarly, DEs immediately after propranolol administration are distinguishable from controls. These differences are maintained 10 min after drug administration. DEs immediately after drug administration were also distinct from DEs 10 min after drug administration, in both the isoproterenol and propranolol cases. Lastly, isoproterenol treated cells were distinguishable from controls even at SNRs of 3 and below. These results indicate that KS-test comparisons of DE widths, upslopes, and downslopes, can accurately distinguish drug-treated DEs from controls, even at low SNRs, and can also distinguish drug-treated DEs based on the time after drug administration.

3.2.4 Discussion and Conclusion

DEC Performance The automated DEC approach allows for quantitative comparison of VSD imaged cardiomyocyte DEs recorded under different drug treatments and time points. In this study, DEC was able to identify significant changes in DE width, upslope, and downslope immediately after treatment with either propranolol or isoproterenol. These changes were sustained when measured 10 min after exposure to each drug. Furthermore, DEC identified changes in DE width between isoproterenol treated cells and controls, even at SNRs of 3 and below.

Propranolol and isoproterenol both act on β-adrenoreceptors in cardiac cells as a β-blocker and β-adrenergic agonist, respectively. β-adrenoreceptors modulate calcium influx during an action potential. Calcium transport from the L-type calcium channels determines the delay before repolarization. This delay determines the width of the DE, also known as the plateau

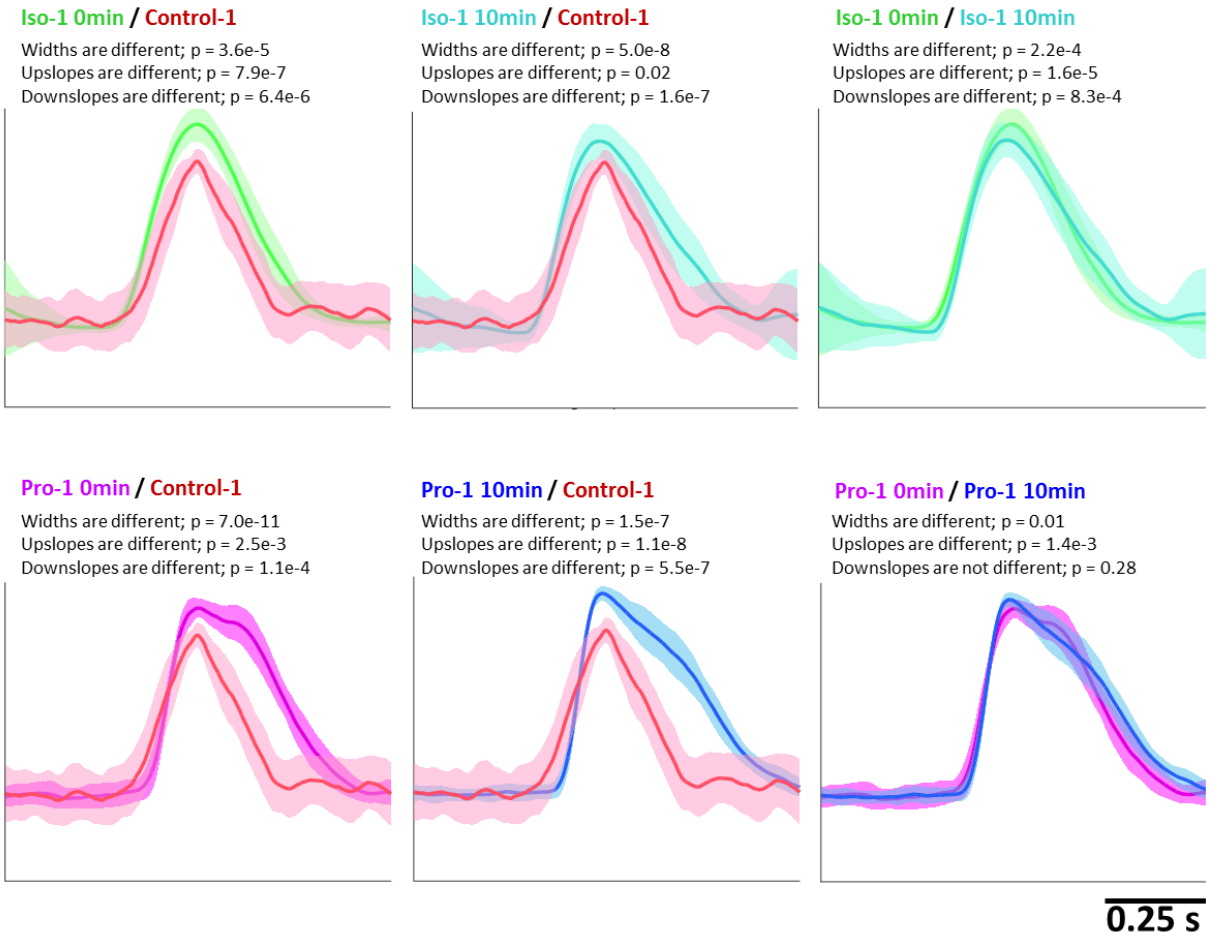

Figure 3.6: DE analysis with respect to drug treatment and time after drug administration. 6 panels are shown. Each panel shows the average DEs of the two specified drug treatments, with standard deviations shaded around the average. The title of each panel identifies the two treatments being compared. The text below the title indicates whether the DE populations were different in width, upslope, and downslope, and provides p-values. The top 3 panels, going from left to right, compare Iso-1 0min (green) to Control-1 (red), Iso-1 10min (light teal) to Control-1, and Iso-1 0min to Iso-1 10min. Both Iso-1 0min and Iso-1 10min were different from Control-1, and were different from each other, in all three parameters (width, upslope, downslope). The bottom 3 panels, going from left to right, compare Pro-1 0min (purple) to Control-1 (red), the Pro-1 10min (blue) to Control-1, and Pro-1 0min to Pro-1 10min. Both Pro-1 0min and Pro-1 10min were different from Control-1 in width, upslope, and downslope. They were also different from each other in width, and upslope.

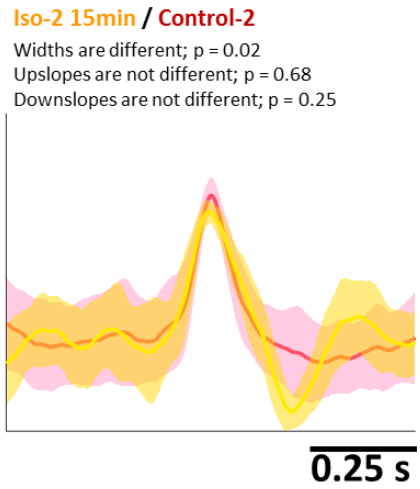

Figure 3.7: DE analysis with respect to drug treatment for the lower SNR data. Iso-2 15min (orange) and Control-2 (red) were different in width, but not in upslope or downslope.

phase. The upslope of a cardiac action potential, primarily driven by fast sodium channels, would not be expected to be affected by β-adrenergic modulation [59]. The downslope may be affected to a lesser degree than the width since the balance of decreasing calcium channel activity and increasing delayed rectifier potassium channel (I_{KS}, I_{KR}, I_{K1}) activity initiate the downslope.

Both isoproterenol- and propranolol-treated cell behavior, quantified by DEC, exhibited significant changes in width and downslope after drug administration. However, the cells also demonstrated an unexpected difference in upslope after drug administration. For the isoproterenol-treated cells, this difference decreased with time (p-value $= 7.9 \times 10^{-7}$ for Iso-1 0min vs Control-1, and p-value $= 0.02$ for Iso-1 10min vs Control-1), and seemed to be trending towards being insignificant. This indicates that the difference in upslope may be due to an initial shock from isoproterenol administration and that the effects may fade as the tissue stabilizes. In contrast, the difference in upslope seemed to increase in time for the propranolol-treated cells (p-value $= 2.5 \times 10^{-3}$ for Pro-1 0min vs Control-1, and p-value $= 1.1 \times 10^{-8}$ for Pro-1 10min vs Control-1). However, both of the p-values are so close to zero, that they're difficult to reliably compare. When low-SNR isoproterenol-treated cell

data (Iso-2 15min) was compared with a low-SNR control (Control-2), only differences in DE widths were detected. Finally, we'd like to note that the DE widths of isoproterenol-treated cells increased with respect to controls, in both low-SNR and high-SNR scenarios. The mechanism of action of isoproterenol (β-adrenergic agonists) would lead us to expect a decrease in the width of the DE. Despite this unexpected, yet consistent, result, DEC still correctly identified and quantified the effect of the drug, emphasizing its impartial approach to DE detection and analysis.

MaD and DEC: MaDEC In order to provide a complete package for VSD imaged cardiomyocyte analysis, the DEC analysis tool was combined with the MaD detector presented in Sec. 2.3.2, for an integrated software called MaDEC. MaDEC was implemented in Matlab and is freely available along with a graphical user interface.

MaDEC can automatically and accurately detect, extract, quantify, and compare VSD-based DEs across drug treatments and time points after drug administration. Unlike other VSD-approaches that use pre-defined waveform features, MaDEC uses a data-driven sample of the entire waveform to detect DEs, resulting in a non-biased DE selection criterion that can accurately detect waveforms at SNRs ≥ 3. Also, unlike other approaches, DE parameters such as width, upslope and downslope, are compared across normalized waveform populations. The method's lack of reliance on absolute fluorescence amplitude ensures that results are not biased by the amount of dye the cells took up, or the cells' exact positions in the imaging plane.

Conclusion DEC is a useful new tool for the study of cardiomyocyte electrophysiology. Combined with the use of VSDs, it allows for non-invasive, image-based, and automated analysis of cardiac DEs. This study demonstrates the ability of this tool to quantify changes in DEs as a function of drug treatment and as a function of time. The software is freely

available and can be easily modified to assess the electrophysiology of other excitable cell populations and data types.

3.3 Conclusion

In this chapter we presented two approaches to electrophysiological data classification. Two main classification problems were discussed, each with a different premise and therefore different solutions that were both nonetheless rooted in a statistical approach.

3.3.1 MUSIC

The first classification problem focused on how to distinguish an unknown number of neurons from multi-sensor extracellular recordings of neuronal APs. Our method of choice focused primarily on feature extraction, and used existing clustering algorithms. We selected neuronal source location as our feature because it is both stable under noisy conditions, and likely to be unique to individual neurons. We localized AP sources using MUSIC and tested the algorithm on both tetrode and heptode data. Because ground truth was not available for these experiments, within- and inter- class analyses were performed to assess classification efficacy.

The MUSIC algorithm applied to the tetrode data used an underlying basic monopole model. All of the 1040 detected APs were successfully localized and 6 distinct clusters were identified in the feature space using EM-BIC. The EM algorithm identified 3% of the APs as outliers. All of the 6 identified classes had different waveform signatures, defined as the signal power ranking across the 4 sensors for each class's average AP. The individual AP waveform signatures were also compared within clusters to determine within-class consistency. Four of the classes achieved above 90% consistency, while two classes were slightly less consistent

at roughly 77%, implying some misclassification. On average the within-class classification consistency was 90%. Overall, this shows that the MUSIC monopole model used with an EM-BIC clustering tool is an accurate classification method that can distinguish several neurons from tetrode data while minimizing the number of APs removed as outliers.

The MUSIC algorithm, this time using an underlying impedance model, was also applied to heptode data. Three data sets were tested and 93%, 50%, and 57% of the detected APs were successfully localized in the three data sets, respectively. This is in contrast to the MUSIC monopole model tested on tetrode data, where all of the detected APs were successfully localized. However, the resulting clusters were very tightly grouped in the feature space, again unlike under the MUSIC monopole model. DBSCAN was then used to cluster the features, and 5-6 classes were identified in each experiment. DBSCAN rejected 4%-7% of the localized APs as outliers. The average AP waveforms for each class were statistically different, as determined by t-tests, for each of the 3 experiments. Furthermore, the classes were very consistent, as can be seen by the minuscule 95% confidence intervals on the average AP waveforms for each class. This implies a very low level of misclassification. Overall, the MUSIC impedance model used with DBSCAN proved to be an accurate tool for sorting spikes from heptode data. Although the method seems to reject more APs as outliers (7%-50% of the APs could not be localized), it also appears to reduce misclassification rates compared with the MUSIC monopole model.

3.3.2 DEC

The second classification problem focused on how to distinguish the DEs of actively beating caridomyocytes imaged using VSD. Being able to distinguish the DEs between different experiments can help in studying the effects of various drugs on cardiomyocyte DE shape and general behavior. Here we applied two β-adrenergic drugs, propranolol and isoproterenol

to cultured hiPS-CMs and imaged the beating cell membranes using VSD. Salient features - upslope, width, and downslope - of the DEs were then compared across two different drug administrations using a K-S test. We called this depolarization event comparison tool DEC.

First, propranolol treated cells imaged immediately after as well as 10 min after drug administration, were compared against controls (not treated with any drug). The same approach was adapted for the isoproterenol treated cells. The drug treated cells' DEs were statistically different from the controls in all three parameters both immediately after as well as 10 min after drug administration for both drugs. The drug treated cells were then compared between drug administration times. The propranolol treated cells immediately after drug administration were statistically different from 10 min after drug administration in upslope and width whereas the isoproterenol treated cells were statistically different in all three parameters.

To further see if this approach can distinguish drug treated DEs from controls when the data SNR is low, we compared a low-SNR control (SNR $\simeq 3$) as well as a low-SNR isoproterenol treated cell culture imaged 15 min after drug administration (SNR $\simeq 0.7$). The average DEs were statistically different in width, but not upslope or downslope.

Overall, DEC was able to accurately distinguish drug treated DEs from controls, even under a low-SNR environment. It was also able to distinguish DEs immediately after drug administration from DEs 10 min after drug administration. This study demonstrates DEC's ability to quantify changes in DEs as a function of drug treatment and as a function of time, which can make it a valuable tool for cardiomyocyte drug assay studies.

Chapter 4

Discovery

The ultimate goal of developing the signal processing techniques outlined in Chapters 2 and 3 is to apply them to electrophysiological data in the process of discovery. This next Chapter outlines exactly this process. Both the multi-sensor extracellular AP detector presented in Sec. 2.3.1 as well as the classification scheme presented in Sec. 3.1.4.2 came together with a few other statistical methods to analyze extracellular heptode data looking at neuronal circuit dynamics as a function of anesthesia depth.

The study presented in this chapter is adapted from Szymanska et al. 2017 [101] and takes a multi-modal approach. Three types of neurological recordings were made simultaneously: electrocorticography (ECoG) collected with intracranial screw electrodes, multi-unit activity (MUA) collected with a heptode, and local field potentials (LFP) also collected with a heptode. Electroencephalography (EEG) and electrocorticography (ECoG) are increasingly used in the clinical and laboratory setting to investigate brain activity. While they provide information from large regions of the brain, little is known regarding how these signals relate to local neuronal activity as measured by microelectrodes. To improve our understanding of how regional signals from the brain surface relate to neuronal activity from deep cortical

layers during varying levels of anesthesia, we recorded ECoG, MUA, and LFP signals during anesthesia induced ECoG burst suppression states at three monotonically increasing anesthesia depths. Based on literature, we hypothesized that increasing anesthesia levels would reduce action potential (AP) firing, reduce the number of active single units, and reduce cross-hemisphere functional connectivity. However, we found that higher anesthesia led to higher AP frequency, no change in the number of active single units, and increased cross-hemisphere functional connectivity. Additionally, all APs were restricted to ECoG bursts, with no APs occurring during suppressed ECoG states. This study demonstrates the effects of deep anesthesia on spontaneous neuronal network dynamics, as well as on the relationship between local (MUA) and global (ECoG and LFP) neurological activity. To our knowledge this is the first study to systematically investigate the relationship between brain surface signals and neuronal activity inside the cortex at various burs suppressed anesthesia depths.

4.1 Background on the Effect of Anesthesia on Neuronal Network Dynamics

Animal and human studies of brain function are often performed under anesthesia, with particular anesthetic agents administered at various doses. However, unlike biochemical effects of anesthetic drugs, which are commonly reported, the effect of anesthesia on neuronal network dynamics remains poorly understood. Furthermore, few studies have focused on deep anesthesia that renders brain activity into a burst suppression pattern on electroencephalogram (EEG) and electrocorticogram (ECoG). Burst suppression activity demonstrates an alternating high voltage pattern interspersed with a suppression pattern, and is commonly achieved under very deep anesthesia and sedation. Little is known about the spiking activity of individual neurons during burst suppression patterns. Similarly, there is a lack of understanding of changes in functional connectivity in the brain during such deep anesthesia.

The majority of studies investigating the effects of anesthesia on neuronal spiking activity concentrate on comparing anesthetized and awake brain states [82, 73, 29]. Although this approach is necessary to determine differences in neuronal function under awake and anesthetized conditions, it does little to elucidate the effects of anesthesia on neuronal networks as the level of anesthesia progressively deepens and the burst suppression ratio (BSR) consequently increases.

Erchova et al. [21] is one of the few recent studies that has investigated neuronal spiking activity under varying levels of anesthesia. The study looked at multi-unit activity (MUA) at intermediate and deep levels of urethane anesthesia in rats, under both spontaneous and stimulus-evoked conditions. The study showed that AP activity was grouped into "bursts" during spontaneous activity, and that the overall spontaneous firing frequency steadily decreased as the level of anesthesia was deepened. However, since only MUA was measured, no inferences could be made regarding the relationship of MUA with global neuronal activity (i.e. functional connectivity), or the relationship of the AP "bursts" to burst suppression activity.

One of the few groups to look at functional connectivity as a function of varying levels of anesthesia, assessed blood oxygenation level dependent (BOLD) signals under light (0.5-1% isoflurane) and moderate (2.9% isoflurane) anesthesia [110]. They found evidence of decreased interhemispheric functional connectivity as isoflurane levels were increased from light to moderate levels. However, the light anesthesia state tested in this study is not traditionally associated with the induction of burst suppression patters, making it resemble the awake vs. anesthetized studies mentioned previously and provides little insight into the effects of increasing anesthesia during a burst suppression state. Furthermore, the BOLD signal is subject to non-neuronal induced fluctuations and its relationship with neuronal activity is often affected by the anesthetics being used [1, 62]. Electrophysiological activity is therefore a more reliable metric for investigating functional connectivity.

EEG as well as ECoG measure regional and global neuronal activity, and are thus good candidates for looking at changes in functional connectivity, including across hemispheres. Although the general consensus is that EEG and ECoG signals originate from the postsynaptic potentials of the apical dendrites of radially aligned pyramidal neurons in the cerebral cortex [77, 58, 93], few studies have looked at the relationship between these signals and the underlying neuronal spiking activity [114, 69, 93]. Most notably, Whittingstall et al. [114], showed that EEG power in the gamma band and phase in the delta band were predictors of MUA. However, none of these studies have investigated this relationship under anesthesia.

This study investigates neuronal firing and its relationship with ECoG signals under varying levels of intermediate and deep anesthesia. To our knowledge, this is the first such study performed. Three experiments were performed on two rats under isoflurane anesthesia. ECoG signals were collected to monitor changes in burst suppression as well as cross-hemisphere functional connectivity as the depth of anesthesia was increased. Similarly, MUA and LFP signals were collected to determine the effects of varying the depth of anesthesia on neuronal firing, and local activity of the neuronal populations. Our hypothesis was that as the anesthesia level was increased, the AP firing frequency would decrease, the cross-hemisphere functional connectivity would decrease, and the number of active single units would decrease; this would be in line with findings from Erchova et al. [21], Noda et al. [73], and Wang et al. [110], respectively. Contrary to our hypothesis, we found that as the level of anesthesia increased, 1) AP firing frequency increased, even as overall AP activity went down, 2) cross-hemisphere functional connectivity, represented by the ECoG-MUA relationship, as well as local population functional connectivity, represented by the LFP-MUA relationship, increased, and 3) the number of active single units was mostly unaffected. A fourth finding we did not anticipate, was that 4) AP activity was strictly relegated to ECoG bursts, with no APs present during suppressed ECoG.

4.2 Methods

4.2.1 Data Collection

The data collection procedure for this work is detailed in Appendix A.3, and was also partially presented in Sec.3.1.4.2. Briefly, two male Wistar rats were implanted with two ECoG screw electrodes for recording over M1 and V1 of the left hemisphere. A burr hole allowing for extracellular heptode recording was made over M1 of the right hemisphere. Figure 4.1 shows a schematic of the locations of the recording sites, with A and C marking the ECoG recording channels and B marking the extracellular recording site. This figure is also provided in Appendix A.3 with a more detailed explanation of the full experimental set up including ground electrodes.

Once the screw electrodes were placed, a motorized head stage was used to lower a heptode into the cortex. The heptode was advanced until a high activity and signal-to-noise ratio (SNR) recording site was found for the experiment. The heptode was used to record both MUA and LFP signals.

4.2.2 Experimental Procedure

The heptode was advanced until a suitable recording site was found, providing adequate SNR and spiking activity, which for these experiments was 1700-1900 μm from the top of the skull. Due to visual limitations inside the burr hole, it was not possible to determine when the heptode reached the surface of the brain; however, assuming a typical thickness of the rat skull of 1 mm, this means the heptode was lowered roughly 700-900 μm into the cortex.

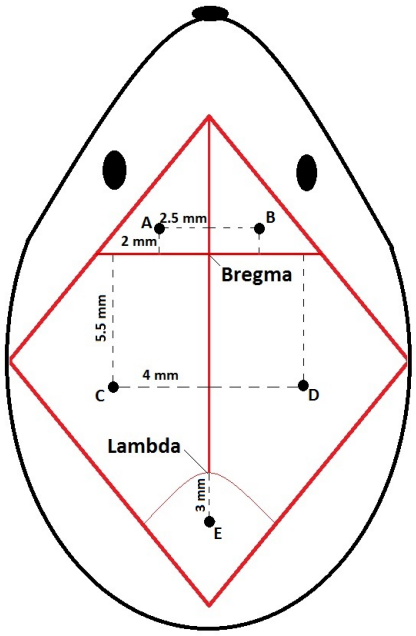

Figure 4.1: Schematic of burr hole locations, as well as ECoG screw electrode and heptode placement. Burr holes are marked as black circles, and placed as specified by the indicated distances from Bregma. The placement is symmetric over the left and right hemispheres. Going from left to right, the top two burr holes are referred to as A and B, the middle two as C and D, and the bottom one as E. Burr holes A, C, D and E are implanted with screw electrodes, whereas B is left open for extracellular recording with a heptode. Screw electrodes A and C are used for ECoG signal recording, screw E is used as the ECoG ground, and screw D is used as the heptode ground.

Once the heptode was in position, 1-2 min of spontaneous neuronal activity were recorded at three monotonically increasing isoflurane anesthesia levels - low, intermediate, and deep. The first measured state represented the lowest level of administered isoflurane necessary to maintaining burst suppression, while the last, deep, level represented the highest level of administered isoflurane prior to a flat ECoG signal. To ensure that isoflurane levels had stabilized prior to recording, a 3-5 min waiting period was observed after isoflurane levels were increased.

The first two experiments were performed on a single animal. After the first recording, another suitable recording site was found 170 μm below the original recording site. This

location was presumably far enough from the initial heptode recording site to be recording from a new population of neurons. These two experiments will be referred to as experiment 1.1 and 1.2, respectively. Experiment 2.1 was performed on the second animal. This animal began going into hypoxia during the deep anesthesia stage, and therefore the last recording was aborted. However, two recordings were made at the low anesthesia stage, one prior to the isoflurane levels being stabilized, and another once the anesthesia level had stabilized. Each recording will be referred to by the experiment number followed by "Iso" and then the percentage of administered isoflurane. For the low anesthesia level recordings in experiment 2.1, the first recording will be referred to as 1 and the second as 2.

4.2.3 Data Processing

Individual APs were detected from the MUA data using the generalized matched filter described Sec. 2.3.1 and Szymanska et al. (2014). The first collected data sets for each experiment (experiment 1.1 Iso 2.0%, experiment 1.2 Iso 1.5%, and experiment 2.1 Iso 1.5% 1) were used as training sets for that experiment. The filter threshold applied to the data ranged from 4.5 - 6.5 median standard deviations above the median filter output [102], depending on the data SNR.

APs were classified using the MUSIC algorithm impedance model for feature extraction (Sec. 3.1.4.2), and DBSCAN for clustering [16] (Appendix B.2). APs identified as outliers by DBSCAN ($\sim$ 15-40 APs depending on the number of APs detected) were manually inspected and re-assigned to existing DBSCAN classes if appropriate. For experiment 1.1 Iso 3.0%, and experiment 1.2 Iso 3.0%, there were not enough APs detected to reliably cluster using the above algorithms. In this case the single units were identified manually from the detected AP waveforms. Only classes of 5 of more APs were retained.

4.2.4 Data Analysis

Filtered data as well as the detected APs were further analyzed to determine various data characteristics as a function of isoflurane level.

4.2.4.1 Burst Suppression Ratio

The burst suppression ratio (BSR) was calculated as one minus the sum of the total duration of all ECoG bursts, divided by the total duration of the ECoG recording.

$$BSR\% = 1 - \frac{1}{T}\sum_{i} Burst_i \tag{4.1}$$

where $Burst_i$ is the duration of burst i, and T is the total duration of the signal.

4.2.4.2 AP Firing Frequency

The AP firing frequency was calculated for each ECoG burst detected in the collected data. An ECoG burst was defined as a period of ECoG activity surrounded by a minimum of 1 s of suppressed ECoG on either side (Fig. 4.2 (**A**)). Within each ECoG burst, there were an average of 1 to 5 AP packets. An AP packet was defined as a period of AP activity surrounded by a minimum of 40 ms of silence on either side (Fig. 4.2 (**B**)). The frequency of spiking activity for each AP packet was calculated as the number of APs detected in the packet divided by the duration of the packet. The frequencies of spiking activity for all packets in a given burst were averaged to determine the AP firing frequency for that burst. All presented AP firing frequencies are calculated per ECoG burst, and then averaged over the ECoG bursts in a given recording.

Table 4.1: Summary metrics for collected data. The first three columns represent the animal used, the corresponding experiment number, as well as the depth of the recording heptode in the cortex (assuming a typical skull width of 1 mm) for each experiment. The following columns provide details about the individual recordings, starting with the percentage isoflurane used for the low, intermediate, and deep anesthesia states, the resulting burst suppression ratios, the average AP firing frequency (± SD), the number of units (or neurons) identified in spike sorting, and the total number of detected APs.

Animal	Expt.	∼ Heptode Depth (μm)	Isoflurane (%)	BSR (%)	Avg. AP Freq. (AP/s ± SD)	Units	Detected APs
1	1.1	700	2.0	68	73±18	5	1184
			2.5	83	135±47	5	443
			3.0	97	243±65	4	62
1	1.2	870	1.5	52	108±32	7	1249
			2.5	88	193±30	5	295
			3.0	96	224±31	5	66
2	2.1	780	1.5	41	125±17	6	5645
			1.5	50	140±15	6	2624
			2.5	80	155±27	6	1107

4.2.4.3 Spike Triggered Average

Spike triggered average analysis was performed on the data in order to investigate the relationship between the MUA and ECoG as well as LFP signals. Once APs were detected for a given data recording, the ECoG signals 300 ms before and after each AP were isolated and averaged, for each ECoG channel. The resulting averaged ECoG signal is the ECoG spike triggered average (ECoG-STA). Similarly, the LFP signals 300 ms before and after each AP were isolated and averaged to calculate the LFP-STA. All STAs are centered on the corresponding AP.

4.3 Results

4.3.1 AP Firing Frequency

In order to investigate the behavior of neuronal firing under different anesthesia levels, all MUA was first detected as described in Sec. 2.3.1. The total numbers of APs detected for each recording are shown in Table 4.1. As expected, the overall spiking activity decreased as the level of administered isoflurane increased, in each of the three experiments. Interestingly, APs were only detected during ECoG bursting activity, throughout all of the experiments presented here (total of $\sim$ 15 min of recording, and 12,675 detected APs). Therefore, the decrease in overall spiking activity followed from the burst suppression ratio increasing. A representative example of this phenomenon is shown in Fig. 4.2. Furthermore, the APs seemed to also exhibit a bursting pattern, within the ECoG bursts, where groups of APs were temporally clustered into 1 to 5 packets, on average, per burst (Fig. 4.2).

Further inspection revealed that the average AP firing frequency (Sec. 4.2.4) within a given ECoG burst was increasing as the level of isoflurane was increasing (Fig. 4.4, Table 4.1). An AP packet was defined as a period of AP activity surrounded by a minimum of 40 ms of silence on either side (Fig. 4.2 (**B**)). The frequency of spiking activity for each AP packet was calculated as the number of APs detected in the packet divided by the duration of the packet. The frequencies of spiking activity for all packets in a given burst were averaged to determine the AP firing frequency for that burst. The AP firing frequencies for each ECoG burst were then averaged for each experiment and level of isoflurane (Fig. 4.4). For experiment 1.1 the AP firing frequency began at 73 $\pm$ 18 APs/s for Iso 2.0%, increased to 135 $\pm$ 47 APs/s for Iso 2.5%, and then increased again to 243 $\pm$ 65 APs/s for Iso 3.0%. The same trend was observed for experiment 1.2 where the firing frequency progressed from 108 $\pm$ 32 APs/s, to 193 $\pm$ 30 APs/s, to 224 $\pm$ 31 APs/s as the level of administered isoflurane was increased from 1.5% to 2.5% to 3.0%, respectively. The same increase, although more

modest, was also observed in experiment 2.1 where the AP firing frequency began at 125 ± 17 APs/s and 140 ± 15 APs/s for Iso 1.5% 1 and Iso 1.5% 2, respectively, and increased to 155 ± 27 APs/s for Iso 2.5%. These results are shown in Fig. 4.4(**A**)

To determine if this increase in AP firing frequency was reflected in the power of the corresponding ECoG bursts, the average ECoG burst power for each experiment and isoflurane level was calculated as the squared sum of the ECoG burst normalized for the duration of the burst. While the increase in AP firing frequency is apparent for each experiment, the ECoG burst power doesn't show any significant trends either across the two ECoG channels, or across the three experiments (Fig. 4.4(**B**)).

Similarly, to investigate if other frequency dependent components of the ECoG burst were coupled with the increasing AP firing frequency, we compared the ECoG spectrograms at different anesthesia levels (Fig. 4.3). This analysis revealed that the frequency of the spectral makeup of the ECoG bursts is minimally affected by increasing the depth of anesthesia. The number and duration of the ECoG bursts were most affected.

These findings suggest that although overall activity decreases as the level of administered isoflurane is increased, when APs do fire, they fire at an increased frequency as the level of isoflurane, and therefore depth of anesthesia, increases. Furthermore, MUA is only present during ECoG bursts, even though the MUA and ECoG activity are recorded cross-hemisphere. The MUA also appears to exhibit a bursting pattern within the bounds of the ECoG bursts, seemingly mimicking bursting behavior but on a smaller timescale.

4.3.2 ECoG-STA Analysis

Spike triggered average analysis was performed on the data in order to investigate the relationship between the MUA and ECoG. Once APs were detected for a given data recording,

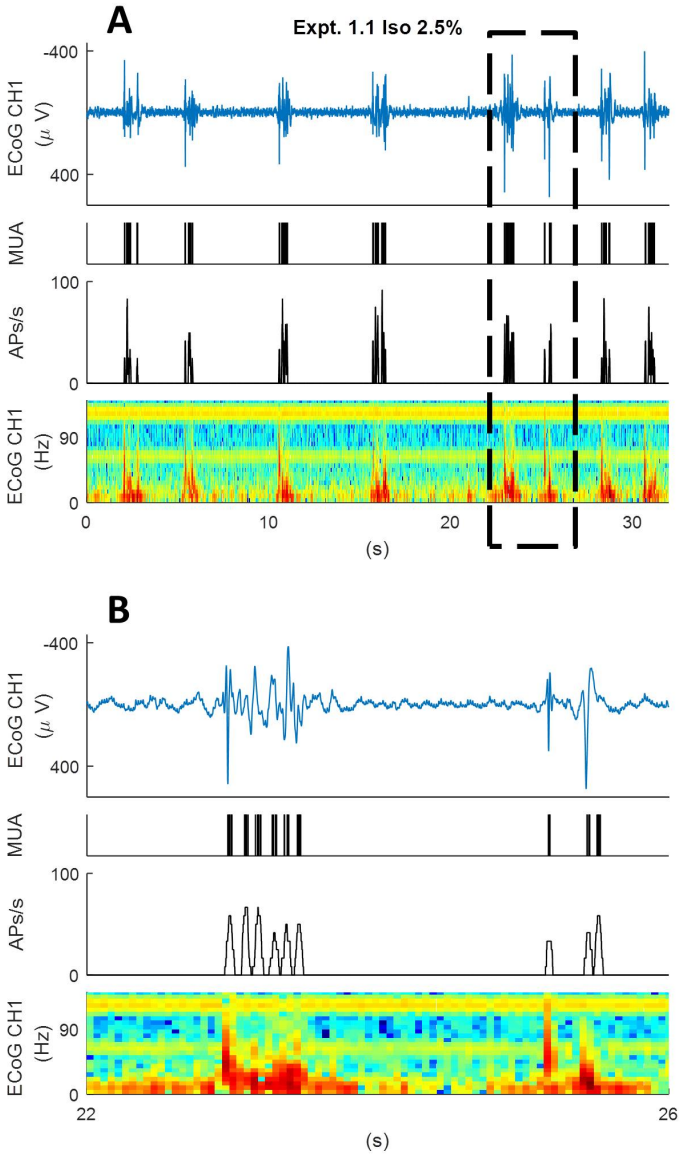

Figure 4.2: Representative ECoG data collected in this study. The ECoG bursts and AP packets are from experiment 1.1 at 2.5% isoflurane. (**A**) *Top* Bursts of ECoG activity on ECoG channel 1 are clearly visible between suppressed ECoG stretches. *Middle Top* Corresponding MUA spike train. The spiking activity is very closely coupled with the ECoG bursts. In fact, the data presented in this study did not contain any APs fired outside of ECoG bursts. *Middle Bottom* Corresponding AP firing frequency (AP/s) calculated using a 40 ms sliding window. *Bottom* Spectrogram of the ECoG data displayed in the top panel. (**B**) A zoom-in of the outlined data in (**A**), showing the activity around 2 ECoG bursts. Looking at the MUA (*Middle Top*) and AP firing frequency (*Middle Bottom*) we can see that the ECoG bursts (*Top*) contain burst-like MUA, where APs seem to be temporally clustered into packets. The first burst contains 6 AP packets, whereas the second bursts contains 3. In this case the individual packets have firing frequencies ranging from 50 to 80 APs/s.

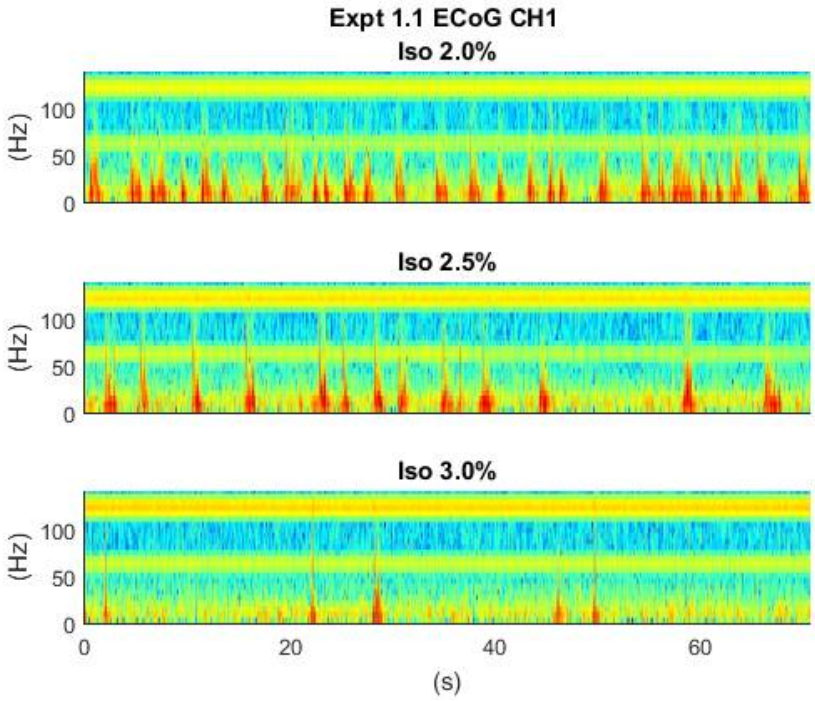

Figure 4.3: Representative ECoG spectrograms as the level of administered isoflurane is increased. The data is from experiment 1.1, ECoG channel 1, at 2.5% isoflurane. As the level of isoflurane is increased, the burst suppression ration increases, and the duration of the bursts decreases. However, the frequency of the spectral makeup of the ECoG bursts is minimally affected by increasing the level of administered isoflurane.

the ECoG signals 300 ms before and after each AP were isolated and averaged, for each ECoG channel. The resulting averaged ECoG signal is the ECoG-STA. (Sec. 4.2.4).

4.3.2.1 General Observations

ECoG-STA analysis was performed on the three experiments across all administered isoflurane levels (Fig. 4.5). The ECoG response, on both channel 1 and channel 2, was significantly above the noise response across all experiments and isoflurane levels, indicating a high degree of functional connectivity between the two hemispheres. This is in line with our previous observation that MUA is highly coupled to ECoG bursts, with no APs present during suppressed ECoG. Across the three experiments the ECoG-STA amplitude tends to be stronger, for all isoflurane levels, for ECoG channel 1 than ECoG channel 2. This is likely due to the

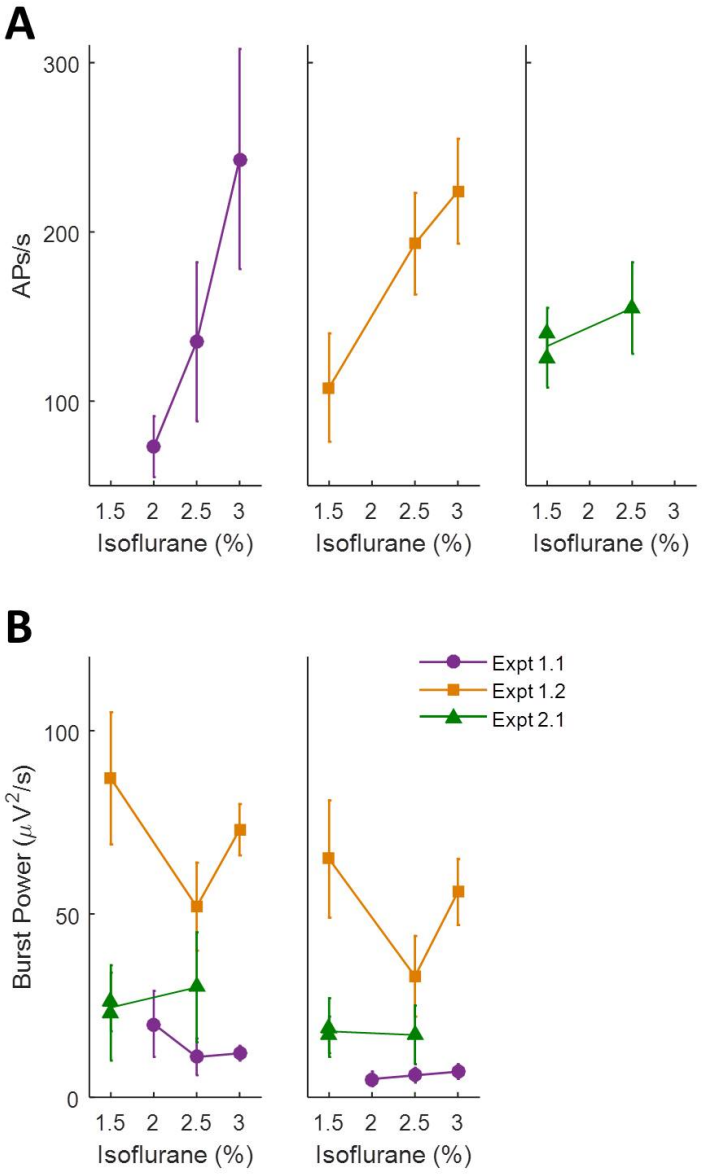

Figure 4.4: AP firing frequency increases as a function of isoflurane, while no obvious trends are visible in the ECoG burst power. (**A**) AP firing frequency as a function of administered isoflurane for experiment 1.1 (*Left*), 1.2 (*Middle*), and 2.1 (*Right*). The error bars represent standard deviations. The firing frequency consistently increases as the level of isoflurane increases. (**B**) ECoG Burst Power as a function of isoflurane level for ECoG channel 1 (*Left*), and ECoG channel 2 (*Middle*). The ECoG burst power was calculated as the squared sum of the ECoG burst normalized for the duration of the burst. The error bars represent standard deviations. The power per ECoG burst doesn't exhibit a steady trend across the three experiments or across the ECoG channels.

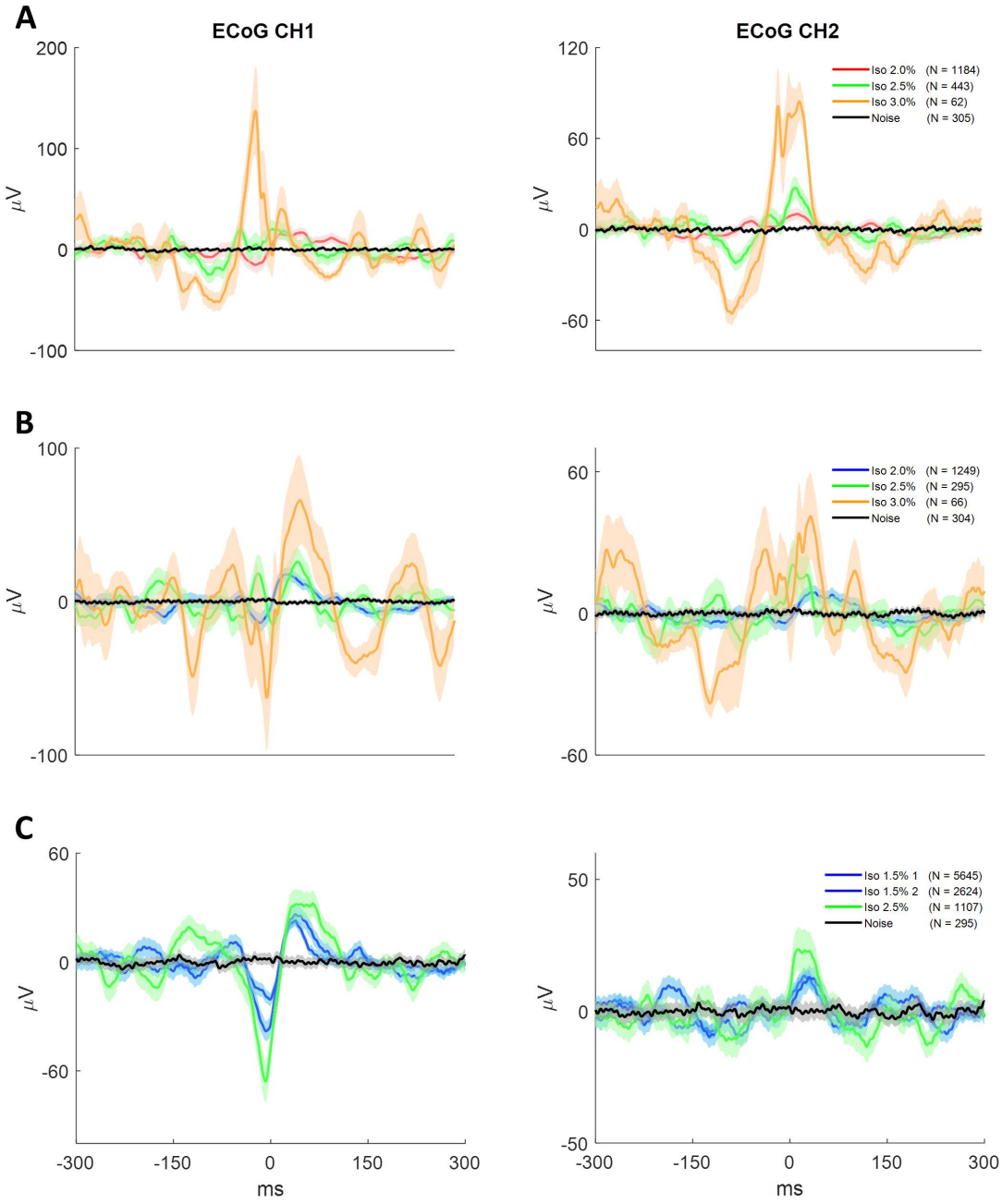

Figure 4.5: As the level of isoflurane is increased, the resulting cross-hemisphere ECoG-STA increases in magnitude for all three experiments. (**A**) (**B**), and (**C**) show ECoG-STAs at varying isoflurane levels for experiment 1.1, 1.2, and 2.1, respectively. The left column shows results for ECoG channel 1, and the right column for ECoG channel 2. The AP was centered at 0 ms for all ECoG-STAs. Shaded regions represent 95% confidence intervals and the number of APs being averaged over is depicted in the three legends. The ECoG-STAs are color coded for isoflurane level, with blue being 1.5%, red 2.0%, green 2.5%, and orange 3.0%.

fact that ECoG channel 1 is over M1 in the frontal lobe, like the MUA recording heptode, whereas ECoG channel 2 is further away over V1. It is therefore expected that channel 2 would exhibit a weaker ECoG-STA response in general, as these two areas are expected to have a lower degree of functional connectivity than the two frontal lobes.

The shape of the ECoG-STA mostly consisted of a negative deflection just before or during the AP and a positive spike immediately after. The exception is experiment 2.1 ECoG channel 2, where the ECoG-STA was monophasic with a positive deflection centered on the AP. Beyond this general shape, it is expected that the phase, width, and specific shape of each ECoG-STA will vary across the three experiments, and in fact this is what we found (Fig. 4.5). The variety of shapes is due to the different neuronal populations under investigation. Although all located in M1, the populations are at slightly different depths (Table 4.1) and from two different animals. They are therefore likely to have slightly different ECoG responses, which manifests in our results as varying ECoG-STA shapes.

Overall, we observed a high amplitude ECoG-STA response, significantly above the noise level, in all experimental cases. This implies a high degree of functional connectivity between the M1 neuronal populations and the cross-hemisphere ECoG activity. Although there were some similarities in the ECoG-STAs, the shapes differed between different neuronal populations as well as different isoflurane levels.

4.3.2.2 ECoG-STA vs. Isoflurane

Fig. 4.5 depicts all ECoG-STAs across the three experiments, ECoG channels, and isoflurane levels, with 95% confidence intervals shown as shaded regions. For all three experiments and across both ECoG channels, the ECoG-STA amplitude increases significantly as the level of isoflurane is increased. This is most pronounced for experiment 1.1 channel 2 (Fig. 4.5 (**A**) *Right*), where the ECoG-STA positive peaks seem to be more aligned than in the other

cases. The trend is less obvious for experiment 1.2 where the shape of the ECoG-STAs and their phase with respect to the AP is less consistent. However the amplitude can still be seen as increasing, with several areas of significance (non-overlapping confidence intervals) in each case.

Furthermore, taking the ECoG-STA root-mean-square (RMS), a proxy for signal power, reveals that even in the more ambiguous cases such as experiment 1.2, the ECoG-STA RMS increases as a function of isoflurane level (Table 4.2). As the level of isoflurane increased from the low to the intermediate state, the increase in ECoG-STA RMS ranged from 158% to 445%. As the level of isoflurane anesthesia was further increased to a deep state for experiments 1.1 and 1.2, the increase in ECoG-STA RMS was even more pronounced ranging from 741% to 1200%.

The ECoG-STA analysis results show that a stronger response, both in terms of ECoG-STA amplitude and ECoG-STA RMS, is produced as isoflurane levels are increased. This implies a stronger degree of coupling between the left hemisphere (represented by the EEG signals), and M1 of the right hemisphere (represented by the MUA), as the level of anesthesia is increased.

4.3.3 LFP-STA

To investigate if this increase in response with increasing levels of anesthesia was also present on the right hemisphere, the same STA analysis was also performed on LFPs (4.2.4). Because each heptode channel is only 40 - 80 μm apart, the 7 recorded LFP channels were nearly identical, therefore only one LFP channel, channel 1, is presented and discussed here. The LFP-STA analysis agreed with the ECoG-STA analysis, showing that as the level of isoflurane increased the response, both in terms of the LFP-STA amplitude and RMS, also increased

Table 4.2: The root-mean-square (RMS) of the ECoG-STA increases as the level of isoflurane increases, for all three experiments. This indicates that the signal is becoming increasingly more coupled with the MUA as the administered isoflurane level is increased.

Experiment	Isoflurane (%)	STA RMS (μV/ms)	
		ECoG CH 1	ECoG CH 2
1.1	2.0	286	147
	2.5	360	312
	3.0	1189	1085
1.2	1.5	239	137
	2.5	348	251
	3.0	1016	683
2.1	1.5 (1)	267	164
	1.5 (2)	425	207
	2.5	673	304

(Fig. 4.6). The LFP-STA shape was more consistent than the ECoG-STA shape. This could be due to the more local nature of the LFP signal.

These results indicate that as the level of isoflurane is increased, global population activity as represented by the ECoG, as well as local population activity as represented by the LFP, become increasingly better coupled with the MUA in M1.

4.3.4 Spike Sorting

To further investigate neuronal firing and behavior under varying levels of anesthesia, detected APs were sorted (Sec. 3.1.4.2) and compared across different isoflurane levels.

Fig. 4.7 shows the resulting average waveforms for each class, across all experiments and levels of isoflurane. For experiment 1.1, 5 units were identified for Iso 2.0% and the same 5 units were also identified for Iso 2.5%. The neuron population persisted at Iso 3.0%, where 4 of the same units were identified, but the fifth was no longer found. Similarly for

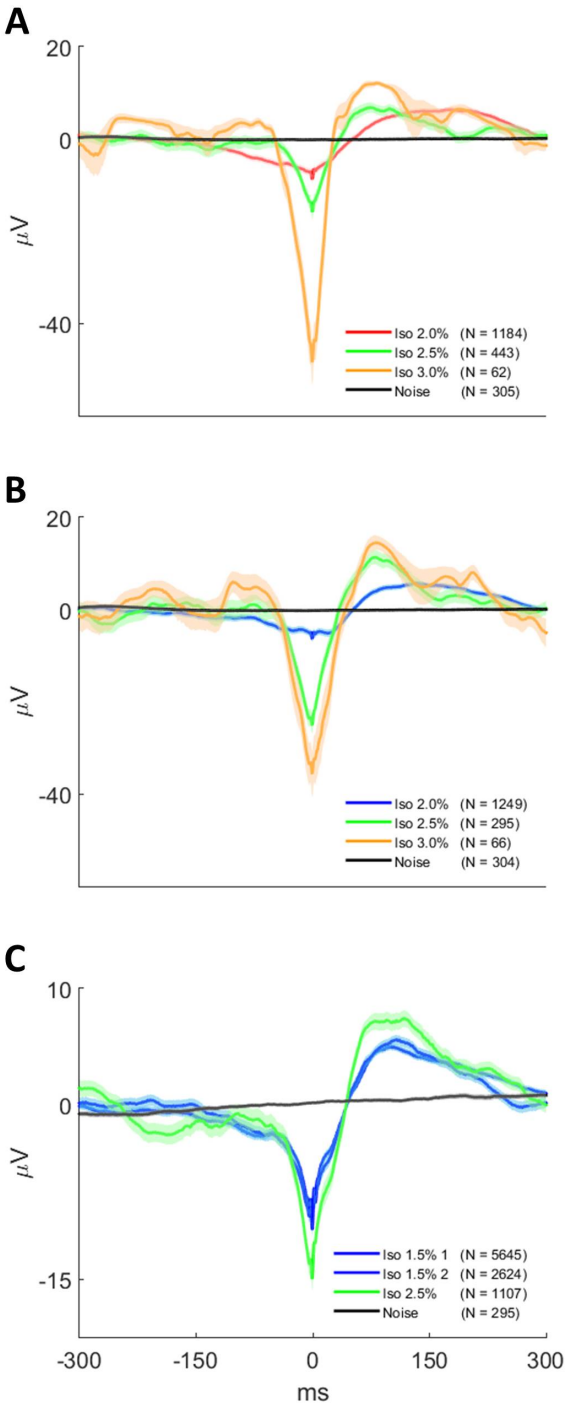

Figure 4.6: As the level of isoflurane is increased, the resulting LFP-STA increases in magnitude for all three experiments, just like the ECoG-STA. **A**, **B**, and **C** show LFP-STAs at varying isoflurane levels for experiment 1.1, 1.2, and 2.1, respectively. The AP was centered at 0 ms for all STAs. Although there are shaded regions representing 95% confidence intervals, they are very small and therefore difficult to discern. The number of APs being averaged over is depicted in the three legends. The STAs are color coded for isoflurane level, with blue being 1.5%, red 2.0%, green 2.5%, and orange 3.0%, just as in Fig. 4.5

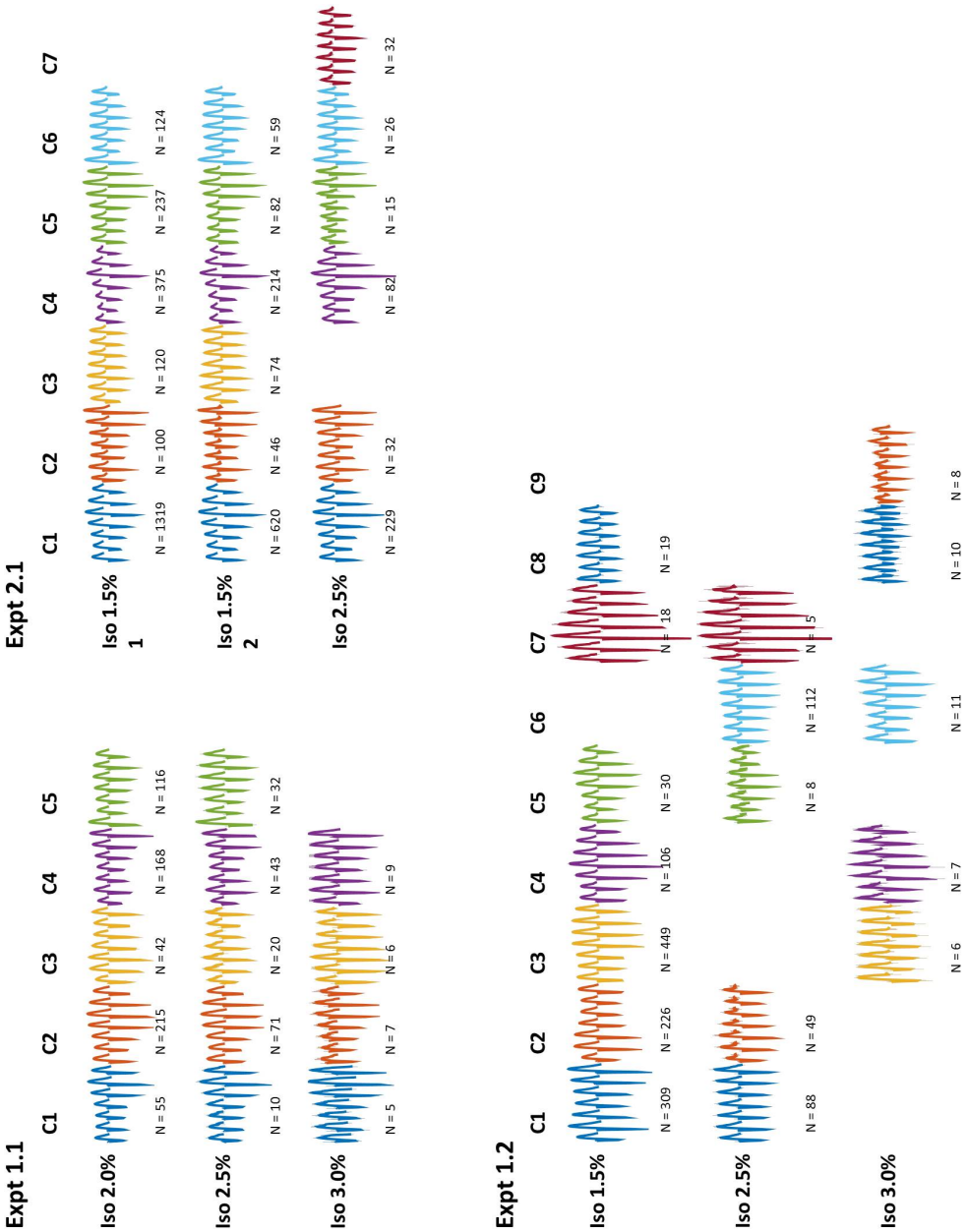

Figure 4.7: The same sub-population of neurons is active as the level of isoflurane is increased. All detected APs were classified for each experiment and level of isoflurane. The 7-channel waveforms for each class are shown, and units appearing across several levels of isoflurane are arranged in columns and color coded. For experiment 1.1 (*Top Left*), 4 units can be traced across all three isoflurane levels, while one unit is lost at Iso 3.0%. Similarly, for experiment 2.1 (*Top Right*), 5 units can be traced across all three isoflurane levels, while one unit is lost and another unit is gained at Iso 2.5%. For experiment 1.2 (*Bottom Left*), no single unit can be traced across all three isoflurane levels, however 8 of the 9 units are present in at least two of the three isoflurane levels, indicating that the same sub-population is active across all three levels of isoflurane.

experiment 2.1, the same 6 units were identified for Iso 1.5% 1 and Iso 1.5% 2. Five of these units were also identified for Iso 2.5%, alongside an additional unit previously not present. For experiment 1.2 a total of 9 units were identified across the three isoflurane levels. Although no single unit persisted through all three levels of isoflurane, like in the other two experiments, 8 of the 9 units recurred in two of the isoflurane levels. Classes 1, 2, 5, and 7 were present for both Iso 1.5% and Iso 2.5%, classes 3, 4, and 8 were present for both Iso 1.5% and Iso 3.0%, whereas class 6 was present for both Iso 2.5% and Iso 3.0%.

Overall these results indicate that the same sub-population of neurons is active as levels of isoflurane are increased during burst suppression. Although a few units disappeared, and a few new ones appeared as the level of anesthesia changed, for the most part the same units could be reconciled across the different levels of isoflurane.

4.4 Discussion and Conclusion

In this study we investigated the effects of varying levels of isoflurane anesthesia on neuronal firing and its relationship with ECoG signals. Our results consist of four main findings.

First, AP firing was tightly coupled with ECoG bursting activity. APs only fired during ECoG bursts, and ECoG bursting always corresponded with AP activity, even though the two signals were recorded on opposite hemispheres. Although in this study there were no examples of APs firing outside of ECoG burst, from prior experience we have observed that there are sometimes APs firing outside of ECoG bursts, but they account for less than 2% of total spiking activity (data not shown).

Second, AP firing frequency increased as the level of anesthesia was increased. The AP activity during ECoG bursts was organized into AP packets, which resembled the ECoG burst but on a smaller time scale. This type of activity is in line with previous findings by Erchova

et al. As the level of anesthesia increased the overall spiking activity decreased, just like in Erchova et al., due to a decreased number of ECoG burst. However, the AP firing frequency within each AP packet increased as the level of anesthesia was increased. This seems to imply that although neurons tended to fire less frequently overall, when activity occurred it was highly synchronous between different neurons. Although Erchova et al. did not look at the AP firing frequency within AP packets, the neuronal activity they recorded across several extracellular electrodes seems to agree with our finding that individual neuronal activity was increasingly more synchronous as the level of anesthesia was increased.

Our third finding was that STA amplitude and power, for both ECoG and LFP signals, increased as the level of isoflurane was increased. These results indicate that ECoG, as well as LFP signals become increasingly better coupled with the MUA as anesthesia is progressively increased. This in turn implies an increase in global as well as local synchrony as the level of anesthesia deepens. This is in line with Hanrahan et al. and Noda et al. who also see evidence of increased AP-LFP phase locking as anesthesia is increased. This phase locking could account for a larger amplitude in the LFP-STAs, although other mechanisms may also be responsible for this phenomenon. The ECoG-STA also increases in amplitude as the anesthesia level is increased, however, this result, implying an increase in functional connectivity, is at odds with Wang et al. whose BOLD signal analysis indicates a decrease in cross-hemisphere functional connectivity as anesthesia is increased. There are several reasons for the discrepancy, the most obvious being that the relationship between BOLD signals and neuronal spiking activity is not well understood, and subject to fluctuations. Although hemodynamic activity is often a proxy to electrophysiological activity, their relationship can change as a result of anesthesia as well as other non-neuronal factors. Another reason may have to do with the scale of the neuronal populations being compared. BOLD signals provide a global measure similar to EEG, which is unlike the very local neuronal activity measure provided by MUA. A study looking at EEG or ECoG signals cross-hemisphere as anesthesia

deepens in a burst suppressed state could identify if synchrony is affected differently at a local / global level vs. a global / global level.

Our last finding was that the number of units seems relatively unchanged as the level of anesthesia was increased. In fact, the same sub-population of neurons seemed to be active throughout the different anesthesia levels. This finding seems to be at odd with Noda et al. who found that the number of active units decreased as the isoflurane level administered to the rat was increased. However, it is worth noting that their study looked at differences between awake and anesthetized rats. It is possible that the number of active units decreases as the isoflurane induces burst suppression, and then remains relatively constant as the anesthesia subsequently deepens. A study monitoring neuronal activity as isoflurane is increased from an awake to a deeply anesthetized state may be able to answer this question.

Although these results are compelling, there are several limitations to this study that we plan to address in our future work. Namely, the small sample size. Our future work will concentrate on increasing the number of experiments. Secondly, it would be beneficial to more precisely determine the layer of the cortex the MUA is being recorded from. The experiments presented here have precise data on the distance traveled by the heptode from the top of the skull, but no definitive data on the exact positioning of the electrode in the brain. To gather this information future experiments should include histology. This would also allow for comparison of the MUA as well as the ECoG-MUA relationship across several cortical layers as a function of increasing anesthesia. Lastly, since this study records MUA from the M1 region only, future studies should also investigate other regions of the brain.

Overall, our four findings suggest that as isoflurane anesthesia is increased during burst suppression, neuronal activity remains bound to the same sub-population of neurons; these neurons' overall activity decreases as the burst suppression ratio goes up. However, when activity does occur it is at a higher firing frequency as the level of anesthesia is increased.

Likewise, the activity is increasingly more synchronous with cross-hemisphere activity, as measured by ECoG, and local population activity as measured by LFP.

Bibliography

[1] D. P. Aksenov, L. Li, M. J. Miller, G. Iordanescu, and A. M. Wyrwicz. Effects of anesthesia on BOLD signal and neuronal activity in the somatosensory cortex. *Journal of Cerebral Blood Flow & Metabolism*, 35(11):1819–1826, 2015.

[2] D. Aur, C. I. Connolly, and M. S. Jog. Computing spike directivity with tetrodes. *Journal of Neuroscience Methods*, 149(1):57–63, 2005.

[3] T. Bányász, Y. Chen-Izu, C. W. Balke, and L. T. Izu. A new approach to the detection and statistical classification of Ca2+ sparks. *Biophysical Journal*, 92(12):4458–4465, 2007.

[4] P. Barthó, H. Hirase, L. Monconduit, M. Zugaro, K. D. Harris, and G. Buzsáki. Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology*, 92(1):600–8, 2004.

[5] M. Bébarová. Advances in patch clamp technique: towards higher quality and quantity. *General Physiology and Biophysics*, pages 131–140, 2012.

[6] R. Benitez and Z. Nenadic. Robust unsupervised detection of action potentials with probabilistic models. *IEEE Transactions on Bio-medical Engineering*, 55(4):1344–1354, 2008.

[7] T. Borghi, R. Gusmeroli, a. S. Spinelli, and G. Baranauskas. A simple method for efficient spike detection in multiunit recordings. *Journal of Neuroscience Methods*, 163(1):176–80, 2007.

[8] M.-A. Bray, N. A. Geisse, and K. K. Parker. Multidimensional detection and analysis of Ca2+ sparks in cardiac myocytes. *Biophysical Journal*, 92(12):4433–4443, 2007.

[9] R. J. Brychta, S. Tuntrakool, M. Appalsamy, N. R. Keller, D. Robertson, R. G. Shiavi, and A. Diedrich. Wavelet methods for spike detection in mouse renal sympathetic nerve activity. *IEEE Transactions on Bio-medical Engineering*, 54(1):82–93, 2007.

[10] G. Buzsáki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5):446–51, 2004.

[11] M. I. Chelaru and M. S. Jog. Spike source localization with tetrodes. *Journal of Neuroscience Methods*, 142(2):305–15, 2005.

[12] R. Cossart, D. Aronov, and R. Yuste. Attractor dynamics of network UP states in the neocortex. *Letters to Nature*, 423:283–288, 2003.

[13] C. Crouzet, R. H. Wilson, A. Bazrafkan, M. H. Farahabadi, D. Lee, J. Alcocer, B. J. Tromberg, B. Choi, and Y. Akbari. Cerebral blood flow is decoupled from blood pressure and linked to EEG bursting after resuscitation from cardiac arrest. *Biomedical Optics Express*, 7(11), 2016.

[14] J. Csicsvari, D. A. Henze, B. Jamieson, K. D. Harris, A. Sirota, P. Barthó, K. D. Wise, and G. Buzsáki. Massively parallel recording of unit and local field potentials with silicon-based electrodes. *Journal of Neurophysiology*, 90(2):1314–23, 2003.

[15] J. Csicsvari, H. Hirase, A. Czurko, and G. Buzsáki. Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron*, 21(1):179–189, 1998.

[16] M. Daszykowski, B. Walczak, and D. L. Massart. Looking for natural patterns in data Part 1: Density-based approach. *Chemometrics and Intelligent Laboratory Systems*, 56:83–92, 2001.

[17] R. R. de Ruyter van Steveninck, G. D. Lewen, S. P. Strong, R. Koberle, and W. Bialek. Reproducibility and variability in neural spike trains. *Science*, 275:1805–1808, 1997.

[18] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2012.

[19] C. Eliasmith and C. H. Anderson. *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. MIT Press, Cambridge, Massachusetts, 2003.

[20] K. L. Ellefsen, B. Settle, I. Parker, and I. F. Smith. An algorithm for automated detection, localization and measurement of local calcium signals from camera-based imaging. *Cell Calcium*, 56(3):147–156, 2014.

[21] I. A. Erchova, M. A. Lebedev, and M. E. Diamond. Somatosensory cortical neuronal population activity across states of anaesthesia. *European Journal of Neuroscience*, 15:744–752, 2002.

[22] M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[23] L. Gaztañaga, F. E. Marchlinski, and B. P. Betensky. Mechanisms of Cardiac Arrhythmias. *Revista Española de Cardiología (English Edition)*, 65(2):174–185, 2012.

[24] I. A. Ghouri, A. Kelly, F. L. Burton, G. L. Smith, and O. J. Kemi. 2-photon excitation fluorescence microscopy enables deeper high-resolution imaging of voltage and Ca2+ in intact mice, rat, and rabbit hearts. *Journal of Biophotonics*, 123(1):112–123, 2013.

[25] E. M. Glaser and M. W. B. Online separation of interleaved neuronal pulse sequences. *Data Acq Proc Biol Med*, 5:137–156, 1968.

[26] S. N. Gozani and J. P. Miller. Optimal discrimination and classification of neuronal action potential waveforms from multiunit, multichannel recordings using software-based linear filters. *IEEE Transactions on Bio-medical Engineering*, 41(4), 1994.

[27] D. S. Greenberg, A. R. Houweling, and J. N. D. Kerr. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature Neuroscience*, 11(7):749–751, 2008.

[28] A. Hajirasooliha. Application of simplified GLRT detectors for spike detection in multi-sensor extracellular recordings. *Masters Thesis, University of California Irvine*, 2013.

[29] S. J. Hanrahan, B. Greger, R. A. Parker, T. Ogura, S. Obara, T. D. Egan, and P. A. House. The effects of propofol on local field potential spectra, action potential firing rate, and their temporal relationship in humans and felines. *Frontiers in Human Neuroscience*, 7:1–11, 2013.

[30] K. Harris, J. Csicsvari, H. Hirase, G. Dragoi, and G. Buzsáki. Organization of cell assemblies in the hippocampus. *Nature*, 424, 2003.

[31] K. Harris, D. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol*, 84(1):401–414, 2000.

[32] K. D. Harris, D. A. Henze, J. Csicsvari, and H. Hirase. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, 84:401–414, 2000.

[33] T. Herron, P. Lee, and J. Jalife. Optical imaging of voltage and calcium in cardiac cells & tissues. *Circulation Research*, 110(4):609–623, 2012.

[34] C. M. Heylman, R. Datta, A. Sobrino, S. George, and E. Gratton. Supervised machine learning for classification of the electrophysiological effects of chronotropic drugs on human induced pluripotent stem cell-derived cardiomyocytes. *PLoS ONE*, 12(10), 2015.

[35] C. M. Heylman, S. Santoso, M. D. Krebs, G. M. Saidel, E. Alsberg, and G. F. Muschler. Modeling and experimental methods to predict oxygen distribution in bone defects following cell transplantation. *Medical & Biological Engineering & Computing*, 52(4):321–30, 2014.

[36] G. R. Holt and C. Koch. Electrical interactions via the extracellular potential near cell bodies. *Journal of Computational Neuroscience*, 6(2):169–84, 1999.

[37] D. Huang and J. Li. The feasibility and limitation of patch-clamp recordings from neonatal rat cardiac ventricular slices. *In Vitro Cellular and Developmental Biology - Animal*, 47:269–272, 2011.

[38] N. Huebsch, P. Loskill, M. A. Mandegar, N. C. Marks, A. S. Sheehan, Z. Ma, A. Mathur, T. N. Nguyen, J. C. Yoo, L. M. Judge, C. I. Spencer, A. C. Chukka, C. R. Russell, P.-L. So, B. R. Conklin, and K. E. Healy. Automated Video-Based Analysis of Contractility and Calcium Flux in Human-Induced Pluripotent Stem-Derived Cardiomyocytes Cultured over Different Spatial Scales. *Tissue Engineering Part C: Methods*, 21(5):467–479, 2014.

[39] D. Humphrey. *Electrophysiological Techniques*. Society for Neuroscience, Atlanta, 1979.

[40] Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste. Synfire chains and cortical songs : Temporal modules of cortical activity. *Science*, 304, 2004.

[41] D. Jäckel, U. Frey, M. Fiscella, F. Franke, and A. Hierlemann. Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of Neurophysiology*, 108(1):334–48, 2012.

[42] M. Jog, C. Connolly, and Y. Kubota. Tetrode technology: advances in implantable hardware, neuroimaging, and data analysis techniques. *Journal of Neuroscience Methods*, 117:141–152, 2002.

[43] S. Joucla, R. Franconville, A. Pippow, P. Kloppenburg, and C. Pouzat. Estimating background-subtracted fluorescence transients in calcium imaging experiments: A quantitative approach. *Cell Calcium*, 54(2):71–85, 2013.

[44] L. Kaestner, A. Scholz, and Q. Tian. Genetically Encoded Ca2+ Indicators in Cardiac Myocytes. *Circulation Research*, pages 1623–1639, 2014.

[45] E. R. Kandel, J. H. Schwartz, and T. M. Jessel. *Principles of Neural Science*. Elsavier, New York, 1991.

[46] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice-Hall, Englewood Cliffs NJ, 1998.

[47] J. N. D. Kerr, D. Greenberg, and F. Helmchen. Imaging input and output of neocortical networks in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):14063–14068, 2005.

[48] S. Kim and J. McNames. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *Journal of Neuroscience Methods*, 165(2):165–174, 2007.

[49] C. Koch. *Biophysics of Computation*. Oxford University Press, New York, 1999.

[50] C. H. T. Kong, C. Soeller, and M. B. Cannell. Increasing sensitivity of Ca2+ spark detection in noisy images by application of a matched-filter object detection algorithm. *Biophysical Journal*, 95(12):6016–6024, 2008.

[51] A. L. Lahti, V. J. Kujala, H. Chapman, A.-P. Koivisto, M. Pekkanen-Mattila, E. Kerkela, J. Hyttinen, K. Kontula, H. Swan, B. R. Conklin, S. Yamanaka, O. Silvennoinen, and K. Aalto-Setala. Model for long QT syndrome type 2 using human iPS cells demonstrates arrhythmogenic characteristics in cell culture. *Disease Models & Mechanisms*, 5:220–230, 2012.

[52] C. W. Lee, H. Dang, and Z. Nenadic. An efficient algorithm for current source localization with tetrodes. *International Conference of the IEEE EMBS*, pages 1282–1285, 2007.

[53] C. W. Lee, C. E. King, S. C. Wu, a. L. Swindlehurst, and Z. Nenadic. Signal source localization with tetrodes: experimental verification. *International Conference of the IEEE EMBS*, 2011:67–70, 2011.

[54] R. Lemon. *Methods for Neuronal Recording in Conscious Animals.* Wiley, New York, 1984.

[55] J. S. Leyton-Mange, R. W. Mills, V. S. Macri, M. Y. Jang, F. N. Butte, P. T. Ellinor, and D. J. Milan. Rapid cellular phenotyping of human pluripotent stem cell-derived cardiomyocytes using a genetically encoded fluorescent voltage sensor. *Stem Cell Reports*, 2(2):163–70, 2014.

[56] X. Lian, J. Zhang, S. M. Azarin, K. Zhu, L. B. Hazeltine, X. Bao, C. Hsiao, T. J. Kamp, and S. P. Palecek. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nature Protocols*, 8(1):162–75, 2013.

[57] X. Liu, X. Yang, and N. Zheng. Automatic extracellular spike detection with piecewise optimal morphological filter. *Neurocomputing*, 79:132–139, 2012.

[58] N. K. Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10):3963–3971, 2003.

[59] J. Ma, L. Guo, S. J. Fiene, B. D. Anson, J. A. Thomson, T. J. Kamp, K. L. Kolaja, B. J. Swanson, and C. T. January. High purity human-induced pluripotent stem cell-derived cardiomyocytes: electrophysiological properties of action potentials and ionic currents. *AJP: Heart and Circulatory Physiology*, 301(5):H2006–H2017, 2011.

[60] W. Q. Malik, J. Schummers, M. Sur, and E. N. Brown. Denoising two-photon calcium imaging data. *PLoS ONE*, 6(6), 2011.

[61] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, CA, 1999.

[62] K. Masamoto and I. Kanno. Anesthesia and the quantitative evaluation of neurovascular coupling. *Journal of Cerebral Blood Flow & Metabolism*, 32(7):1233–1247, 2012.

[63] F. Mechler and J. D. Victor. Dipole characterization of single neurons from their extracellular action potentials. *Journal of Computational Neuroscience*, 32(1):73–100, 2012.

[64] F. Mechler, J. D. Victor, I. Ohiorhenuan, A. M. Schmid, and Q. Hu. Three-dimensional localization of neurons in cortical tetrode recordings. *Journal of Neurophysiology*, 106(2):828–48, 2011.

[65] N. Mordwinkin, P. Burridge, and J. Wu. A review of human pluripotent stem cell-derived cardiomyocytes for high-throughput drug discovery, cardiotoxicity screening and publication standards. *Journal of Cardiovascular Translational Research*, 6(1):22–30, 2013.

[66] L. Moreaux and G. Laurent. Estimating firing rates from calcium signals in locust projection neurons in vivo. *Frontiers in Neural Circuits*, 1(2):1–13, 2007.

[67] M. Moya, D. Tran, and S. C. George. An integrated in vitro model of perfused tumor and cardiac tissue. *Stem Cell Research & Therapy*, 4 Suppl 1(Suppl 1):S15, 2013.

[68] E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–760, 2009.

[69] S. Musall, V. V. Pföstl, A. Rauch, N. K. Logothetis, and K. Whittingstall. Effects of neural synchrony on surface EEG. *Cerebral Cortex*, 24(4):1045–1053, 2014.

[70] K. Nakae, Y. Ikegaya, T. Ishikawa, S. Oba, H. Urakubo, M. Koyama, and S. Ishii. A statistical method of identifying interactions in neuronglia systems based on functional multicell Ca2+ imaging. *PLoS Computational Biology*, 10(11):e1003949, 2014.

[71] Z. Nenadic and J. W. Burdick. Spike detection using the continuous wavelet transform. *IEEE Transactions on Bio-medical Engineering*, 52(1):74–87, 2005.

[72] Z. Nenadic and J. W. Burdick. A control algorithm for autonomous optimization of extracellular recordings. *IEEE Transactions on Bio-medical Engineering*, 53(5):941–55, 2006.

[73] T. Noda and H. Takahashi. Anesthetic effects of isoflurane on the tonotopic map and neuronal population activity in the rat auditory cortex. *European Journal of Neuroscience*, 42(6):2298–2311, 2015.

[74] H. Norimoto, M. Mizunuma, D. Ishikawa, N. Matsuki, and Y. Ikegaya. Muscarinic receptor activation disrupts hippocampal sharp wave-ripples. *Brain Research*, 1461:1–9, 2012.

[75] M. Oheim, M. van 't Hoff, A. Feltz, A. Zamaleeva, J.-M. Mallet, and M. Collot. New red-fluorescent calcium indicators for optogenetics, photoactivation and multi-color imaging. *Biochimica et Biophysica Acta*, 1843(10):2284–306, 2014.

[76] K. Oweiss. *Multiresolution Analysis of Multichannel Neural Recordings in the Context of Signal Detection, Estimation, Classification and Noise Suppression*. PhD thesis, U. of Michigan, 2002.

[77] C. J. D. Pomfrett. The EEG during anaesthesia. *Current Anaesthesia and Critical Care*, 9(3):117–122, 1998.

[78] C. Pouzat and G. Laurent. Locust data available online at http://www.biomedicale.univ-paris5.fr/SpikeOMatic/Data.html. 2005.

[79] C. Pouzat, O. Mazor, and G. Laurent. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *Journal of Neuroscience Methods*, 122(1):43–57, 2002.

[80] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16(8):1661–87, 2004.

[81] P. Ramdya, B. Reiter, and F. Engert. Reverse correlation of rapid calcium signals in the zebrafish optic tectum in vivo. *Journal of Neuroscience Methods*, 157(2):230–237, 2006.

[82] R. L. Rennaker, H. L. Carey, S. E. Anderson, A. M. Sloan, and M. P. Kilgard. Anesthesia suppresses nonsynchronous responses to repetitive broadband stimuli. *Neuroscience*, 145(1):357–369, 2007.

[83] M. Rizk and P. D. Wolf. Optimizing the automatic selection of spike detection thresholds using a multiple of the noise level. *Medical & Biological Engineering & Computing*, 47(9):955–66, 2009.

[84] M. Ronzhina, V. Cmiel, O. Janoušek, J. Koláová, M. Nováková, P. Babula, and I. Provazník. Application of the optical method in experimental cardiology: action potential and intracellular calcium concentration measurement. *Physiological Research / Academia Scientiarum Bohemoslovaca*, 62(2):125–37, 2013.

[85] M. Sahani. *Latent Variable Models for Neural Data Analysis*. PhD thesis, California Institute of Technology, 1999.

[86] T. Sasaki, R. Kimura, M. Tsukamoto, N. Matsuki, and Y. Ikegaya. Integrative spike dynamics of rat CA1 neurons: a multineuronal imaging study. *The Journal of Physiology*, 574(1):195–208, 2006.

[87] T. Sasaki, N. Matsuki, and Y. Ikegaya. Metastability of active CA3 networks. *The Journal of Neuroscience*, 27(3):517–528, 2007.

[88] T. Sasaki, N. Takahashi, N. Matsuki, and Y. Ikegaya. Fast and accurate detection of action potentials from somatic calcium fluctuations. *Journal of Neurophysiology*, 100(3):1668–76, 2008.

[89] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34:276–280, 1986.

[90] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

[91] V. Shalchyan, W. Jensen, and D. Farina. Spike detection and clustering with unsupervised wavelet optimization in extracellular neural recordings. *IEEE Transactions on Bio-medical Engineering*, 59(9):2576–85, 2012.

[92] D. Smetters, A. Majewska, and R. Yuste. Detecting action potentials in neuronal populations with calcium imaging. *Methods: A Companion to Methods in Enzymology*, 18(2):215–221, 1999.

[93] A. C. Snyder and M. A. Smith. Stimulus-dependent spiking relationships with the EEG. *Journal of Neurophysiology*, 114(3):1468–1482, 2015.

[94] Z. Somogyvári, L. Zalányi, I. Ulbert, and P. Erdi. Model-based source localization of extracellular action potentials. *Journal of Neuroscience Methods*, 147(2):126–37, 2005.

[95] C. I. Spencer, S. Baba, K. Nakamura, E. A. Hua, M. A. Sears, C.-C. Fu, J. Zhang, S. Balijepalli, K. Tomoda, Y. Hayashi, P. Lizarraga, J. Wojciak, M. M. Scheinman, K. Aalto-Setälä, J. C. Makielski, C. T. January, K. E. Healy, T. J. Kamp, S. Yamanaka, and B. R. Conklin. Calcium Transients Closely Reflect Prolonged Action Potentials in iPSC Models of Inherited Cardiac Arrhythmia. *Stem Cell Reports*, 3:269–281, 2014.

[96] M. E. Spira and A. Hai. Multi-electrode array technologies for neuroscience and cardiology. *Nature Nanotechnology*, 8(2):83–94, 2013.

[97] E. M. Steele and D. S. Steele. Automated detection and analysis of Ca2+ sparks in x-y image stacks using a thresholding algorithm implemented within the open-source image analysis platform ImageJ. *Biophysical Journal*, 106(3):566–576, 2014.

[98] A. Szymanska, M. Doty, K. V. Scannell, and Z. Nenadic. A supervised multi-sensor matched filter for the detection of extracellular action potentials. *International Conference of the IEEE EMBS*, pages 5996–9, 2014.

[99] A. Szymanska, A. Hajirasooliha, and Z. Nenadic. Source location as a feature for the classification of multi-sensor extracellular action potentials. *International Conference of the IEEE EMBS NER*, 2013.

[100] A. Szymanska and Z. Nenadic. Wavelet-approximated generalized matched filter for the detection of multisensor extracellular action potentials. *International Conference of the IEEE EMBS NER*, 2013.

[101] A. F. Szymanska, M. Alcala Alvarez, M. Farahabadi, A. Bazrafkan, Z. Nenadic, and Y. Akbari. Effects of increasing anesthesia on spontaneous neuronal activity during burst suppression in M1 of rat cortex. *PLoS ONE*, Submitted, 2017.

[102] A. F. Szymanska, M. Doty, K. V. Scannell, and Z. Nenadic. A supervised multi-sensor matched filter for the detection of extracellular action potentials. *International Conference of the IEEE EMBS*, pages 3–6, 2014.

[103] A. F. Szymanska, C. Heylman, R. Datta, E. Gratton, and Z. Nenadic. Automated detection and analysis of depolarization events in human cardiomyocytes using madec. *Computers in Biology and Medicine*, 75, 2016.

[104] A. F. Szymanska, C. Kobayashi, H. Norimoto, Y. Ikegaya, and Z. Nenadic. Accurate detection of low signal-to-noise ratio neuronal calcium transient waves using a matched filter. *Journal of Nueroscience Methods*, 259:1–12, 2016.

[105] N. Takahashi, K. Kitamura, N. Matsuo, M. Mayford, M. Kano, N. Matsuki, and Y. Ikegaya. Locally synchronized synaptic inputs. *Science*, 335(6066):353–6, 2012.

[106] N. Takahashi, T. Sasaki, A. Usami, N. Matsuki, and Y. Ikegaya. Watching neuronal circuit dynamics through functional multineuron calcium imaging (fMCI). *Neuroscience Research*, 58(3):219–25, 2007.

[107] P. H. Thakur, H. Lu, S. S. Hsiao, and K. O. Johnson. Automated optimal detection and classification of neural action potentials in extra-cellular recordings. *Journal of Neuroscience Methods*, 162(1-2):364–76, 2007.

[108] S. Ueno, M. Tsukamoto, T. Hirano, K. Kikuchi, M. K. Yamada, N. Nishiyama, T. Nagano, N. Matsuki, and Y. Ikegaya. Mossy fiber Zn2+ spillover modulates heterosynaptic N-methyl-D-aspartate receptor activity in hippocampal CA3 circuits. *Journal of Cell Biology*, 158(2):215–220, 2002.

[109] A. Vallmitjana, M. Barriga, Z. Nenadic, A. Llach, E. Alvarez-Lacalle, L. Hove-Madsen, and R. Benitez. Identification of intracellular calcium dynamics in stimulated cardiomyocytes. *International Conference of the IEEE EMBS*, pages 68–71, 2010.

[110] K. Wang, M. P. A. Van Meer, K. Van Der Marel, A. Van Der Toorn, L. Xu, Y. Liu, M. A. Viergever, T. Jiang, and R. M. Dijkhuizen. Temporal scaling properties and spatial synchronization of spontaneous blood oxygenation level-dependent (BOLD) signal fluctuations in rat sensorimotor network at different levels of isoflurane anesthesia. *NMR in Biomedicine*, 24(1):61–67, 2011.

[111] L. Wang, T. Chen, X. Zhou, Q. Huang, and C. Jin. Atomic force microscopy observation of lipopolysaccharide-induced cardiomyocyte cytoskeleton reorganization. *Micron*, 51:48–53, 2013.

[112] F. v. Wegner, M. Both, and R. H. A. Fink. Automated detection of elementary calcium release events using the á trous wavelet transform. *Biophysical Journal*, 90(6):2151–2163, 2006.

[113] J. N. Weiss, A. Garfinkel, H. S. Karagueuzian, P.-S. Chen, and Z. Qu. Early afterdepolarizations and cardiac arrhythmias. *Heart Rhythm*, 7(12):1891–1899, 2010.

[114] K. Whittingstall and N. K. Logothetis. Article frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron*, 64(2):281–289, 2009.

[115] M. Wilson and B. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058, 1993.

[116] S. Wu, A. Swindlehurst, P. Wang, and Z. Nenadic. Projection versus prewhitening for EEG interference suppression. *IEEE Transactions on Bio-medical Engineering*, 59(5):1329–1338, 2012.

[117] S.-C. Wu, A. L. Swindlehurst, and Z. Nenadic. A novel framework for feature extraction in multi-sensor action potential sorting. *Journal of Neuroscience Methods*, 253:262–271, 2015.

[118] S. C. Wu, A. L. Swindlehurst, P. T. Wang, and Z. Nenadic. Projection versus prewhitening for EEG interference suppression. *IEEE Transactions on Bio-medical Engineering*, 59(5):1329–38, 2012.

[119] E. Yaksi and R. W. Friedrich. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. *Nature Methods*, 3(5):377–383, 2006.

[120] P. Yan, C. D. Acker, W.-L. Zhou, P. Lee, C. Bollensdorff, A. Negrean, J. Lotti, L. Sacconi, S. D. Antic, P. Kohl, H. D. Mansvelder, F. S. Pavone, and L. M. Loew. Palette of fluorinated voltage-sensitive hemicyanine dyes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):20443–8, 2012.

[121] X. Yang and S. Shamma. A totally automated system for the detection and classification of neural spikes. *IEEE Transactions on Bio-medical Engineering*, 35(10):806–816, 1988.

[122] Y. Yuan, C. Yang, and J. Si. The M-Sorter: an automatic and robust spike detection and classification system. *Journal of Neuroscience Methods*, 210(2):281–90, 2012.

[123] N. Zeevi-Levin, J. Itskovitz-Eldor, and O. Binah. Cardiomyocytes derived from human pluripotent stem cells for drug screening. *Pharmacology & Therapeutics*, 134(2):180–8, 2012.

Appendix A

Data Acquisition Methods

This appendix contains detailed procedures for the collection of data used throughout this work, with relevant citations included.

A.1 VSD Imaging of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes

This section provides details on the data collection methods for the voltage sensitive dye (VSD) imaged human induced pluripotent stem cell-derived cardiomyocytes (hiPS-CMs) used in Chapters 2.3.2 and 3.2. The data used here was also employed in Szymanska et al. [103] as well as Heylman, et. al [34], and the detailed methods are also provided in these articles.

A.1.1 Human Induced Pluripotent Stem Cell-Derived Cardiomyocyte (hiPS-CM) Culture and Differentiation

hiPS-CMs were prepared for interrogation per the protocol previously described by Heylman, et. al [34]. Briefly, wtc11 hiPS cells were differentiated into cardiomyocytes using a serum-free defined medium protocol [56]. Cells began spontaneously beating on approximately Days 12-15 , and were stained with VSD and imaged on Day 33.

A.1.2 Voltage-Sensitive Dye Staining and Drug Exposure

Culture medium was replaced with fresh medium containing 1μM Di-4-ANE(F)PPTEA (purchased from Leslie Loew, University of Connecticut) and incubated for 15 min at 37^{o}C. Cells were rinsed with RPMI/B-27 (+) insulin one time and then allowed to recover for at least 2 hours prior to imaging. After staining with VSD, cells were qualitatively confirmed to still be spontaneously beating before addition of drugs. Medium was then replaced with fresh medium containing either $10^{-5}\mu$M propranolol (SIGMA, P0884) or $10^{-7}\mu$M isoproterenol (SIGMA, I6504). Data was collected immediately after addition of drugs (less than 60 sec of exposure) and again 10 min or 15 min after addition to ensure complete exposure. Control images were captured from VSD stained cultures not treated with either drug.

A.1.3 Two-Photon Microscopy

A Zeiss LSM 710 microscope (Carl Zeiss, Jena, Germany) with a 40X water immersion objective (C-Apochromat 40X/1.20 W Korr M27) was used for all measurements. VSD was excited by an 850nm light produced by a titanium:sapphire Mai Tai laser (Spectra-Physics, Mountain View, CA). Excitation light was separated from emission signal with a 760nm dichroic. VSD fluorescence was collected in the 489 - 645 nm range. Line scan

mode with 128 pixels per line and a 1.58 μs pixel dwell time was used to acquire temporal VSD depolarization data. Given a 1.67 kHz sampling rate, the total scan time per line was 600 μs. Each measurement consisted of 100,000 line scan repeats (total scan time 60 s). The Zen software package (Zeiss, Jena, Germany) was used to control all microscope components and acquisition processes. Brightfield images were used to identify clusters of spontaneously beating cardiomyocytes. The system was then switched to line scan mode with the parameters specified above. Line scan data were acquired along a line that was manually drawn across cell membranes. After completion of data acquisition, the system was switched back to brightfield mode to confirm that the cells were still spontaneously beating.

A.1.4 Data Pre-Processing

SimFCS commercial software developed in the Laboratory of Fluorescence Dynamics (LFD, University of California, Irvine) was used to analyze raw fluorescence data. A Gaussian tracking and correction algorithm (Supplemental Fig. A.1) was used to compensate for motion artifacts resulting from the spontaneous beating of cell clusters. Fluorescence intensity along each corrected cell membrane trace was then extracted. Finally, a custom Matlab script that fit and subtracted a biexponential function from the resultant data was used to remove photobleaching artifacts (Supplemental Fig. A.2).

A.1.5 Manual Identification of Depolarization Event

Intensity traces, X, for all drug and control conditions, as described in Sec. A.1.2, were derived as described in Sec. A.2.5, and plotted in Matlab. Three trained human analysts then independently identified depolarization event (DE) peak times from each trace.

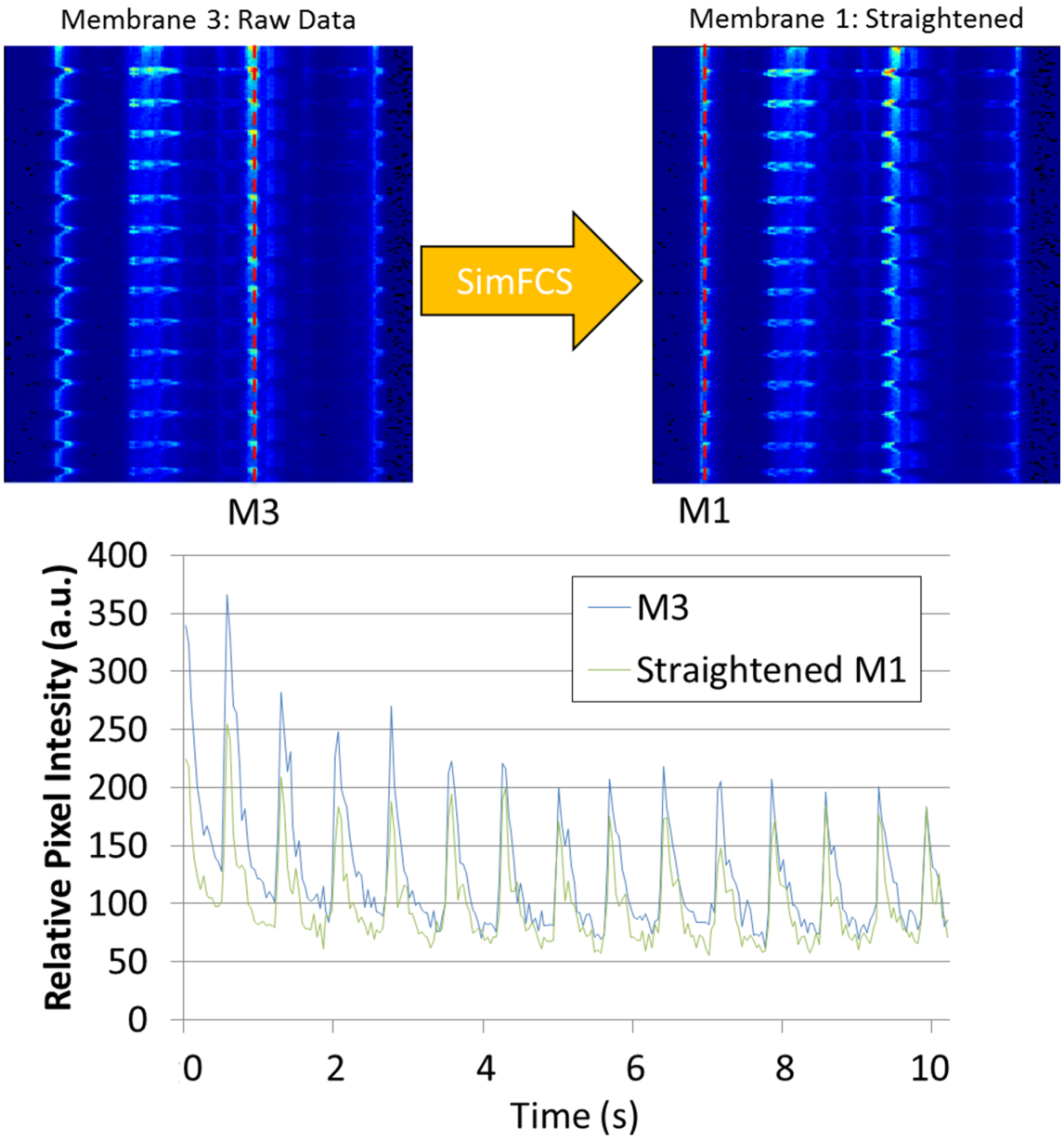

Figure A.1: Motion Artifact Correction. Motion artifact resulting from the spontaneous beating of cell clusters was compensated for in data pre-processing using a Gaussian tracking and correction algorithm. Quantification of membrane 3 depolarization peaks using pre-corrected raw data correlates well with membrane 1 depolarization peaks quantified using data corrected with the Gaussian tracking algorithm.

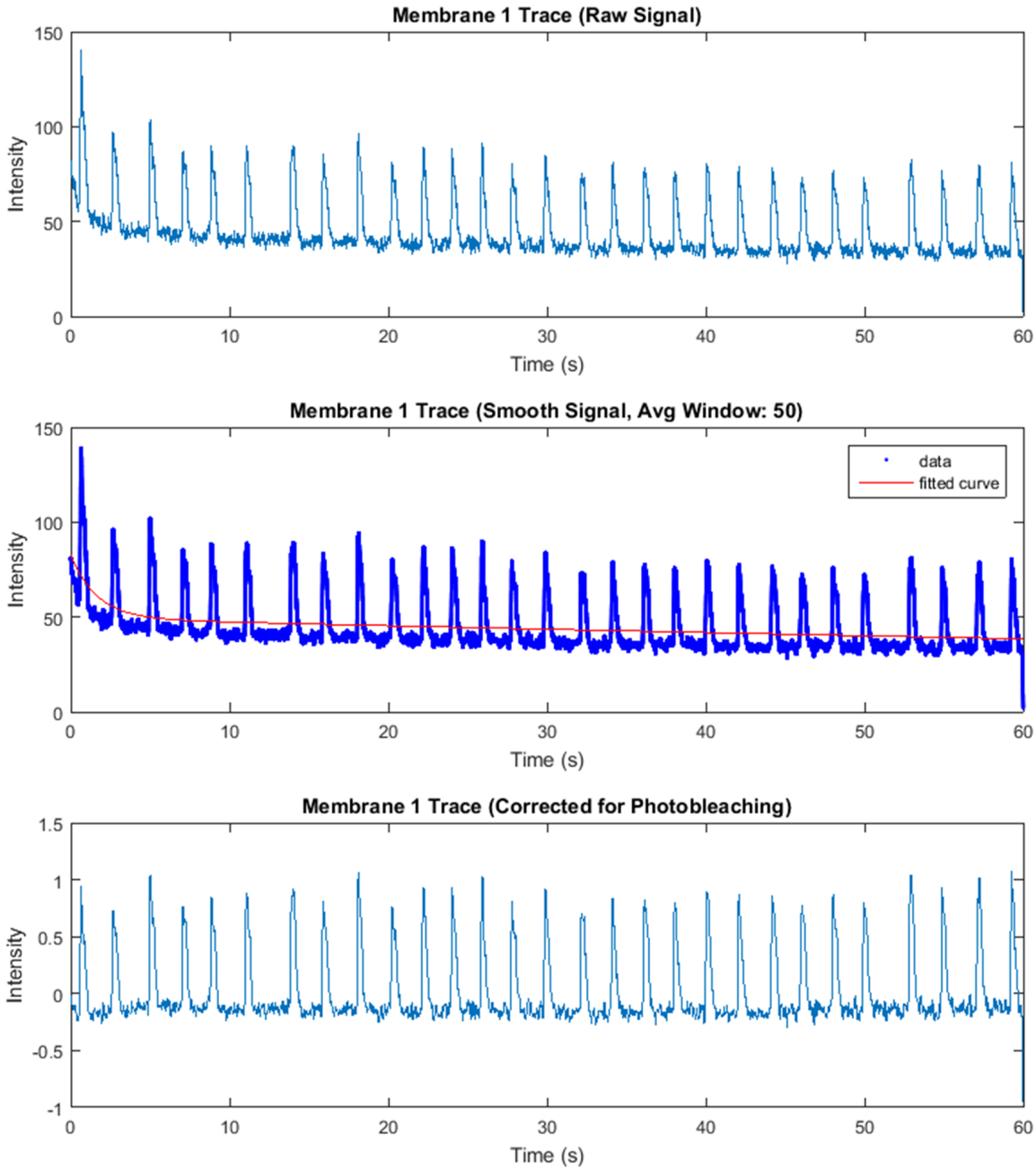

Figure A.2: Photobleaching Correction. Photobleaching was accounted for by fitting a biexponential ($y = ae^{bx} + ce^{dx}$) and subtracting from the signal. (*Top*) Raw signal collected by the instrument. (*Middle*) Overlay of second order exponential fit and raw signal. (*Bottom*) Resultant signal after subtracting the exponential fit and normalizing to baseline fuorescence of the membrane (F_o).

A.2 Functional Multi-unit Calcium Imaging of Neuronal Somas and Dendritic Spines

This section provides details on the data collection methods for functional multi-unit calcium imaging (fMCI) of neuronal somas as well as dendritic spines. The data described here was used in Chapter 2.3.3, and also employed in Szymanska et al. [104], where these details are also provided.

A.2.1 Tissue Preparation

All animals used in this study were treated according to The University of Tokyo guidelines for the care and use of laboratory animals. All performed experiments were approved by the experiment ethics committee at the University of Tokyo, approval numbers: P24-5 and P24-8.

Acute Slices Acute slices were prepared as described in Ueno et al. 2002 [108, 74]. Briefly, 400 μm horizontal slices of the hippocampus from 3 week old C57BL/6J mice were prepared using a vibratome in ice-cold oxygenated cutting solution. The slices then rested in oxygenated artificial cerebrospinal fluid (aCSF) at room temperature for 1.5 hours prior to imaging. For more details on the preparation and solutions used please see [108, 74].

Ex Vivo Cultures In order to facilitate the simultaneous visibility of many dendritic spines during fMCI, organotypic slice cultures were used in this study. *Ex vivo* rat hippocampal slice cultures were prepared as described in Takahashi et al. [105] from 7 day old Wistar/ST rats. Briefly, 300 μm entorhinal-hippocampal organotypic slices were cut using a vibratome, placed on Omnipore membrane filters (JHWP02500; Millipore, Bedford, Mas-

sachusetts, USA), and incubated (5% CO_2, 37^{o}C) in culture medium (50% minimal essential medium, 25% Hanks' balanced salt solution, 25% horse serum, antibiotics) for 12-19 days prior to imaging. The medium was changed every 3.5 days. For more details please refer to [105].

A.2.2 Dye Loading

Simultaneous Somatic fMCI and Patch-Clamp *Ex vivo* slice cultures were transferred into a dish (35-mm diameter) containing 2 ml of the dye solution and were incubated for 1-h in a humidified incubator at 35°C under 5% CO_2. The dye solution was aCSF containing 10 μl of 0.1% Oregon Green BAPTA1-AM (OGB1) dissolved in DMSO, 2 μl of 10% Pluronic F-127/DMSO and 2 μl of 5% Cremophor EL/DMSO. After being washed, the cultured slices were incubated at 35°C for 40 min and were mounted in a recording chamber.

Somatic fMCI Acute slices were loaded locally with OGB1 dissolved in DMSO containing 10% Pluronic F-127 to yield a concentration of 2 mM. Immediately before use, this solution was 10× diluted with aCSF and loaded into pipettes (3 - 5 MΩ). The tip of the pipette was inserted into the CA1 stratum pyramidale, and a pressure was applied using a 10-ml syringe pressurizer (50–60 hPa for 5 min).

Dendritic Spine fMCI CA3 pyramidal neurons were selected for spine imaging from *ex vivo* slice cultures. Selected neurons were voltage-clamped at -30 mV (MultiClamp 700B amplifier and a Digidata 1440A digitizer controlled by pCLAMP 10.4 software) and loaded with a Fluo-4 solution (97.3 $CsMeSO_4$, 42.7 CsCl, 10 HEPES, 10 phosphocreatine, 4 MgATP, 0.3 NaGTP, and 0.2 Fluo-4, all in mM). The voltage clamp facilitated channel currents mediated by NMDA receptors and calcium-permeable AMPA receptors.

A.2.3 Patch-Clamp Recording

CA3 pyramidal neurons selected for simultaneous somatic fMCI and patch-clamp recordings were voltage-clamped at 0 mV (Axopatch 700B amplifier) using a borosilicate glass pipettes (4 - 9 MΩ) filled with aCSF, and recorded from at a sampling frequency of 20kHz (MultiClamp 700B amplifier and a Digidata 1440A digitizer controlled by pCLAMP 10.4 software).

A.2.4 Optical Recording

Fluorophores were excited at 488 nm with a laser diode (HPU50101PFS, FITEL, Tokyo, Japan) and visualized using a 507-nm long-pass emission filter. Videos were taken at 10 Hz for the simultaneous somatic fMCI and patch-clamp data and at 50 Hz for the somatic fMCI and dendritic spine fMCI data using a Nipkow-disk confocal microscope (CSU-X1; Yokogawa Electric, Tokyo, Japan), and a cooled EM-CCD camera (iXon DU897, Andor, Belfast, UK). A 16x objective was used for somatic fMCI data and a 60x objective was used for the dendritic spine fMCI data (CFI75LWD16xW and CFI75LWD60xW Nikon, Tokyo, Japan).

One data trace was acquired for simultaneous somatic fMCI and patch-clamp (fMCI-PC) recordings. The recording lasted 300-s and measured spontaneous activity from a CA3 pyramidal neuron from the *ex vivo* slice cultures. Three somatic fMCI data sets were acquired measuring spontaneous activity from CA1 pyramidal neurons from the acute slices. Recordings for each data set lasted 60-s and contained from 75 to 190 regions of interest (ROIs). Five dendritic spine fMCI data sets were acquired measuring spontaneous synaptic inputs from CA3 pyramidal neurons from the *ex vivo* slice cultures. Recordings ranged from 54-s to 111-s and contained from 157 to 313 ROIs. Only spines located within 200 μm from the soma were monitored to avoid the space clamp problem.

A.2.5 Optical Data Pre-Processing

ROIs for all somatic fMCI and dendritic spine fMCI data sets were identified manually using custom software in Microsoft Visual Basic [40]. A single analyst (Analyst 5 from Table 2.7) identified all ROIs for the somatic fMCI data sets, and another analyst (Analyst 6 from Table 2.7) identified all ROIs for the dendritic spine fMCI data sets. These common sets of ROIs were used in all further analysis. The diameter of each ROI was tailored to the visible size of the neuron soma or dendritic spine being identified. For fMCI-PC data, the ROI was drawn to exclude the micropipette.

The average fluorescence for each ROI was then calculated and used to determine the change in fluorescence, $\Delta F/F = (F_1 - F_0)/F_0$, where F_1 is the fluorescence intensity at any time point, and F_0 is the average baseline fluorescence intensity 1-s before and after F_1. This normalization with a 2-s epoch around the target frame is needed to compensate for photobleaching. The resulting fluorescence intensity trace, X, for each ROI was then used for further analysis.

A.2.6 Manual CE Identification

Somatic and dendritic spine fluorescence intensity traces, X, derived as described in Sec. A.2.5, were independently plotted in Matlab. Six trained human analysts then identified CE peak times from the traces. On average, the human analysts spent 24 working hours identifying calcium events (CEs) in a given data set, where each data set contained between 75 and 313 ROIs (179 ROIs on average). CEs were not manually identified for fMCI-PC data.

A.3 Extracellular and ECoG Recording from Anesthetized Rats

This section provides details on the *in vivo* data collected from rats used to study the effects of anesthesia depth on spontaneous neuronal network dynamics presented in Chapter 4. This work has been submitted for publication [101] and the detailed methods will also soon be provided in this article.

A.3.1 Animal Selection

Two male Wistar rats, both weighing 305 g were included in the study, and a total of three experiments were performed. The animals were part of a separate study which called for the implantation of intracranial screw electrodes [13]. The current study was performed during the screw electrode implantation surgery. All animals were treated in accordance to University of California, Irvine guidelines and the experimental procedures were approved by the local Institutional Animal Care and Use Committee.

A.3.2 Screw Electrode Implantation Surgery

Rats were initially anesthetized in a seal-tight induction chamber connected to a vaporizer supplying 5.0% isoflurane gas carried by 100% oxygen gas at 4L/min for 5 minutes. The rats were then removed from the chamber and fitted with an adult animal facemask/nosecone supplying 1.5-2.5% isoflurane. Adequate depth of anesthesia for the surgery was determined by absence of corneal, forelimb, and/or hindlimb withdrawal reflex, and was continuously monitored throughout the experiment. The rat's heads were then shaved and they were

secured in a stereotactic frame. A 2.5cm rostral-caudal incision was made in a sterile fashion after cleaning the skin with iodine and allowing the iodine to dry.

Once the cranium was exposed, five 1.5-2 mm burr holes were drilled through the skull using a mini hand drill. Bregma and lambda, two landmarks on the skull, were located to measure the points at which burr holes would be drilled. Two frontal lobe burr holes were drilled 2 mm anterior to bregma, and each 2.5 mm to the left and right of bregma, corresponding to the M1 region of the motor cortex. These locations will be referred to as A and B respectively. Two parietal burr holes were drilled 5.5 mm posterior to bregma, and each 4 mm lateral on the left and right of bregma, corresponding to the V1 region of the visual cortex. These locations will be referred to as C and D respectively. The last burr hole, referred to as E, was drilled 3 mm posterior to lambda over the cerebellum. Figure A.3 shows a schematic of the locations of each burr hole.

Screw electrodes (Plastics One. Catalog number E363-20) were implanted into all burr holes, except for burr hole B, which was used for MUA and LFP recording. Screw electrodes A and C were used to collect ECoG signals, with screw electrode E serving as the ground. Note that this type of recording is also often referred to as intracranial EEG. Screw electrode D served as the reference for the MUA and LFP recordings.

A.3.3 Microelectrode Placement

Once screw electrodes A, C, D, and E were placed, the dura under burr hole B was removed and a motorized head stage (Thomas Recording Tetrode Mini Matrix) housing a 7-channel extracellular microelectrode (Thomas Recording), or heptode, was positioned over burr hole B, immediately above the surface of the burr hole. The motorized head stage was used to lower the heptode into the cortex, until a high activity and signal-to-noise ratio (SNR) recording site was found for the experiment.

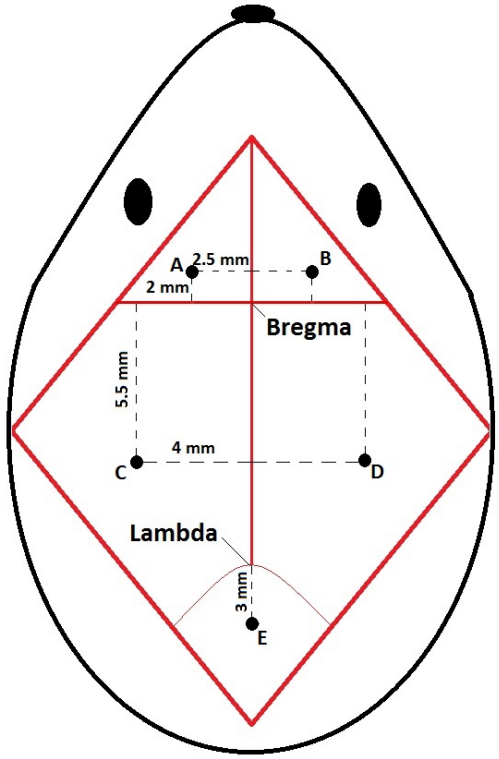

Figure A.3: Schematic of burr hole locations, as well as ECoG screw electrode and heptode placement. Burr holes are marked as black circles, and placed as specified by the indicated distances from Bregma. The placement is symmetric over the left and right hemispheres. Going from left to right, the top two burr holes are referred to as A and B, the middle two as C and D, and the bottom one as E. Burr holes A, C, D and E are implanted with screw electrodes, whereas B is left open for extracellular recording with a heptode. Screw electrodes A and C are used for ECoG signal recording, screw E is used as the ECoG ground, and screw D is used as the heptode ground.

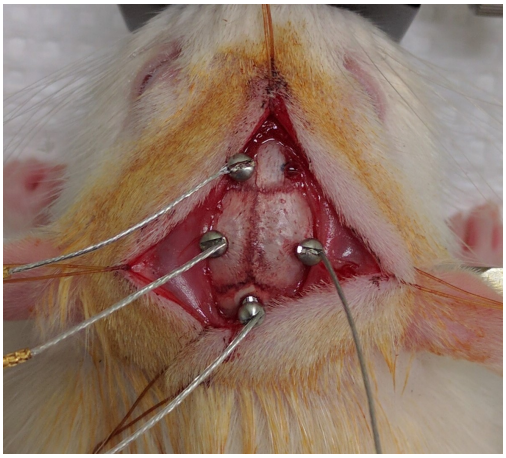

Figure A.4: Image of implantation surgery, with all burr holes drilled and screw electrodes placed.

A.3.4 Recording

ECoG data was acquired from implanted screw electrodes at 1526 Hz using a PZ2 preamplifier (Tucker-Davis Technologies Inc., Alachua, FL) and an RZ5D data acquisition system (TDT).

The heptode was used to record both MUA and LFP signals. The signals collected by the head stage were passed to a pre-amplifier (TDT, Medusa), and then to the data acquisition system (TDT, RX7). The data acquisition system then band-pass filtered the data at 300-3000 Hz and stored both the filtered and raw data. The filtered data was used to extract MUA, whereas the raw data was used for LFP signals.

All collected data sets, including ECoG, MUA, and LFP signals, were notch filtered to remove 60 Hz noise. The ECoG data was then band pass filtered at 2.5-110 Hz, and the LFP signals were band pass filtered at 2.5-300 Hz. Because the heptode recording channels were only 40-80 μm apart, the LFP recordings from each channel were essentially identical. Therefore one LFP channel, channel 1, was used for all subsequent analysis.

A.3.5 Experimental Procedure

Once all ECoG and extracellular recording electrodes were in place, the automated motorized microdrive on the heptode headstage was used to lower the heptode into the burr hole and then into the cortex. The heptode was advanced until a suitable recording site was found, providing adequate SNR and spiking activity, which for these experiments was 1700-1900 μm from the top of the skull. Due to visual limitations inside the burr hole, it was not possible to determine when the heptode reached the surface of the brain; however, assuming a typical thickness of the rat skull of 1 mm, this means the heptode was lowered roughly 700-900 μm into the cortex.

Once the heptode was in position, 1-2 min of spontaneous neuronal activity were recorded at three consecutively increasing isoflurane anesthesia levels - low, intermediate, and deep. The first measured state represented the lowest level of administered isoflurane necessary to maintaining burst suppression, while the last, deep, level represented the highest level of administered isoflurane prior to a flat ECoG signal. To ensure that isoflurane levels had stabilized prior to recording, a 3-5 min waiting period was observed after isoflurane levels were increased, prior to the next recording.

The first two experiments were performed on a single animal. After the first recording, another suitable recording site was found 170 μm below the original recording site. This location was presumably far enough from the initial heptode recording site to be recording from a new population of neurons. These two experiments will be referred to as experiment 1.1 and 1.2, respectively. Experiment 2.1 was performed on the second animal. This animal began going into hypoxia during the deep anesthesia stage, and therefore the last recording was aborted. However, two recordings were made at the low anesthesia stage, one prior to the isoflurane levels being stabilized, and another once the anesthesia level had stabilized. Each recording will be referred to by the experiment number followed by "Iso" and then the

percentage of administered isoflurane. For the low anesthesia level recordings in experiment 2.1, the first recording will be referred to as 1 and the second as 2.

A.3.6 Post-experiment

Once the experiments were completed, the level of isoflurane was decreased to 1.5-2.5%, and the extracellular probe was removed from burr hole B and replaced by a screw electrode, in preparation for a subsequent study [13]. All screw electrodes were connected to an adapter, and fixed in place using dental cement. The rats were then injected with buprenorphine 0.02 mg/kg subcutaneously for analgesia 5 min prior to stopping isoflurane and allowing recovery from anesthesia.

Appendix B

Clustering Methods

B.1 EM-BIC Method

The classification algorithm used on the MUSIC-derived source locations in Sec. 3.1.4.1 is based on Expectation-Maximization (EM) and Bayes' information criterion (BIC). Given a number of expected clusters, the EM algorithm determines the best cluster parameters. In the case of source location features, these parameters include each cluster's center and variance. Note that this means each feature point is assigned a specific cluster. Optimal cluster assignments are determined using the EM algorithm for several cluster models. The best model is then selected based on the BIC.

B.1.1 Expectation Maximization

The central premise of the EM algorithm is to expand the data R into a more complete data vector Y by introducing what's called the missing data vector, B

$$Y = \begin{bmatrix} R \\ B \end{bmatrix}$$

where R is the vector of all 3-D localized APs detected from the data. The missing data vector B should specify the cluster assignments of each feature, or location, $r_i \in R$. The algorithm assumes there are K clusters being modeled, each with a Gaussian distribution, $\mathcal{G}_k$, $k \in [1, K]$. The outliers, on the other hand, have a uniform distribution $\mathcal{U}$. We can then express the components of the missing data vector B as

$$b_i = \begin{cases} [1, 0, \cdots, 0] & \text{if} \quad r_i \in \mathcal{U} \quad \text{(outlier)} \\ [0, 1, \cdots, 0] & \text{if} \quad r_i \in \mathcal{G}_1 \quad \text{(cluster 1)} \\ & \ddots \\ [0, 0, \cdots, 1] & \text{if} \quad r_i \in \mathcal{G}_K \quad \text{(cluster K)} \end{cases}$$

where $b_i \in \mathbb{R}^{1 \times K}$ takes on one of these values with probability τ_k, $\sum_{k=0}^{K} \tau_k = 1$, τ_0 represents the probability of b_i denoting an outlier, and $\tau = [\tau_0, \cdots, \tau_K]$.

The ultimate parameters being solved for are each cluster's center and variance, which can be expressed as

$$\theta = \begin{bmatrix} A \\ \upsilon \end{bmatrix}$$

where A represents each clusters center in Cartesian coordinates, and υ represents each cluster's variance.

The EM is designed to maximize the complete data log-likelihood

$$L(Y|\theta) = \sum_{i=1}^{N} \ln f(y_i|\theta)$$

where N is the total number of feature points, and f represents probability density function (PDF). Because the missing data B is unknown, the complete data log-likelihood cannot be evaluated and must be estimated instead. To do this we first estimate the complete log-likelihood function

$$Q(\theta) = E[L(Y|\theta)|R, \theta^j] \tag{B.1}$$

where E is the expectation operator, and θ^j is a previous estimate of the parameter θ. Given that B is a set of independent random variables we can express Eq. B.1 as

$$Q(\theta) = \sum_{i=1}^{N} \sum_{\text{all } b_i} \ln f(y_i|\theta) \times P(b_i|r_i, \theta^j) \tag{B.2}$$

The probability density $f(y_i|\theta)$ can be derived fairly easily. Remember that y_i is a point in the full data vector Y which consists of both the extracted features, R, and the missing data, B. The probability of b_i being in cluster k is τ_k and the probability of r_i being in cluster k is $\mathcal{G}_k(r_i, \theta)$, for $k \in [1, K]$, and $\mathcal{U}(r_i)$, for $k = 0$.

$$f(y_i|\theta) = \begin{cases} \tau_0 \mathcal{U}(r_i) & \text{if } b_i = [1, 0, \cdots, 0] \quad \text{outlier} \\ \tau_k \mathcal{G}_k(r_i, \theta) & \text{else} \quad \text{depending on the cluster } k \end{cases} \tag{B.3}$$

We can now express the prior probability as

$$P(b_i|r_i, \theta^j) = \frac{f(r_i|b_i, \theta^j) P(b_i|\theta^j)}{\sum_{\text{all } b_i} f(r_i|b_i, \theta^j) P(b_i|\theta^j)} \tag{B.4}$$

where the individual terms of Eq. B.4 are

$$f(r_i|b_i,\theta^j)P(b_i|\theta^j) = \begin{cases} \tau_0^j \mathcal{U}(r_i) & \text{if } b_i = [1,0,\cdots,0] \quad \text{outlier} \\ \tau_k^j \mathcal{G}_k(r_i,\theta^j) & \text{else} \quad \text{depending on the class } k \end{cases} \tag{B.5}$$

and τ_k^j represents the previous estimate of the probability of b_i denoting a given cluster k.

Given this we can finally calculate the complete log-likelihood function $Q(\theta)$ in Eq. B.1. This is the expectation phase of the algorithm. The EM algorithm then applies the following iteration:

- Initialization: Select initial parameter estimates θ^0, and τ^0 as your current parameters. Set the iteration limit to J.
- Expectation Phase: Estimate the complete log-likelihood function using your current parameters, θ^j in Eq. B.2, and τ^j in Eq. B.5
- Maximization Phase: Evaluate

  $$\theta^{j+1} = \arg\max_{\theta} Q(\theta) \tag{B.6}$$

  Note that this maximization will also implicitly determine $\tau^{j+1} = P(b_i|\theta^{j+1})$.
- Iterate: If $j+1 < J$ and $|\theta^{j+1} - \theta^j|^2 > \varepsilon$, where ε is a preset tolerance parameter, set $j = j+1$ and repeat the Expectation Phase. Otherwise, the optimal parameters have been evaluated.

B.1.2 Bayes' Information Criterion

The EM algorithm determines the optimal cluster parameters given a set number of clusters K. However, the initial number of clusters is not generally know. The calculation is therefore

repeated for M different cluster models. Each model, $\mathcal{M}_i$, $i \in M$, is then evaluated based on the BIC

$$\begin{aligned} \mathrm{BIC}(\mathcal{M}_i) &= L(R|\mathcal{M}_i) - \tfrac{\nu_i}{2} \ln N \\ L(R|\mathcal{M}_i) &\cong L(Y|\theta_i^*) \end{aligned} \tag{B.7}$$

where ν_i is the number of independent parameters in the model, N is the total number of features in the data set, and the maximum log-likelihood was approximated by the EM algorithm using the optimal parameters θ_i^* for each model $\mathcal{M}_i$. The optimal cluster model is then selected as the one that maximizes the BIC

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} \mathrm{BIC}(\mathcal{M}) \tag{B.8}$$

B.2 NP-DBSCAN

The clustering algorithm used on the MUSIC-derived source locations in Sec. 3.1.4.2 and Chapter 4, uses the density-based spatial clustering of applications with noise (DBSCAN) approach first introduced by Ester et al. [22]. The algorithm was expanded on by Daszykowski et al. [16] to better detect natural patterns (NP) and this is the ultimate algorithm used here. The algorithm is available online, and the full algorithm details are presented in Daszykowski et al. 2001. For the benefit of the reader, this section we will briefly explain the basis of the NP-DBSCAN algorithm.

DBSCAN is a single-scan approach that assumes density-based clusters. The main idea of this approach is that the object density within a given neighborhood has to exceed a certain specified threshold in order to form a cluster. The threshold is usually set as a minimum number of objects, k, present within a neighborhood radius, ϵ.

The approach identifies three types of objects in the data set. Core objects are objects that contain the minimum number of neighbors, k, within their neighborhood radius, ϵ. Border objects are within the neighborhood of a core object, but do not contain k neighbors in their own neighborhood. Lastly, outliers do not contain k neighbors in their neighborhood, and are not within the neighborhood of a core object. Core and border objects form clusters within the data set.

Unlike with other approaches, including EM, there is no need to specify the number of expected clusters prior to classification using this method. Instead, the main parameters that need to be optimized are the minimum number of neighbors, k, and the neighborhood radius, ϵ. Daszykowski et al. developed an approach for optimization of the neighborhood radius that relies solely on the data size and the minimum number of neighbors, k. Given an $m \times n$ feature set, F, where m is the number of features and n is the dimensionality of the feature space, the NP approach simulates a set of m objects in the n-dimensional feature space with the same range as the original features set F. However, the simulated objects are uniformly distributed within the range of the experimental feature set. The distance from each "artificial" object i to its kth nearest neighbor is then calculated, and ϵ is set as the 95% quantile distance.

For the purposes of MUSIC based spike sorting, m was the number of localized APs, and $n = 3$ was the dimension of the features space. We set the minimum number of neighbors to $k = 12$ and used a Euclidean distance measure.