2014

A METHOD FOR NON-INVASIVE, AUTOMATED BEHAVIOR CLASSIFICATION IN MICE, USING PIEZOELECTRIC PRESSURE SENSORS

Steven R. Gooch

University of Kentucky, s.ryan.gooch@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Steven R. Gooch, Student

Dr. Kevin Donohue, Major Professor

Dr. Caicheng Lu, Director of Graduate Studies

A METHOD FOR NON-INVASIVE, AUTOMATED BEHAVIOR CLASSIFICATION IN MICE, USING PIEZOELECTRIC PRESSURE SENSORS

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Electrical Engineering in the College of Engineering
at the University of Kentucky

By

Steven Ryan Gooch

Lexington, Kentucky

Director: Dr. Caicheng Lu, Professor of Electrical Engineering

Lexington, Kentucky

2014

ABSTRACT

A METHOD FOR NON-INVASIVE, AUTOMATED BEHAVIOR CLASSIFICATION IN MICE, USING PIEZOELECTRIC PRESSURE SENSORS

While all mammals sleep, the functions and implications of sleep are not well understood, and are a strong area of investigation in the research community. Mice are utilized in many sleep studies, with electroencephalography (EEG) signals widely used for data acquisition and analysis. However, since EEG electrodes must be surgically implanted in the mice, the method is high cost and time intensive. This work presents an extension of a previously researched high throughput, low cost, non-invasive method for mouse behavior detection and classification. A novel hierarchical classifier is presented that classifies behavior states including NREM and REM sleep, as well as active behavior states, using data acquired from a Signal Solutions (Lexington, KY) piezoelectric cage floor system. The NREM/REM classification system presented an 81% agreement with human EEG scorers, indicating a useful, high throughput alternative to the widely used EEG acquisition method.

KEYWORDS: Classification, Mice, Behavior Analysis, Linear Discriminant Analysis, Noninvasive Data Acquisition

Steven Ryan Gooch

May 5, 2014

A METHOD FOR NON-INVASIVE, AUTOMATED BEHAVIOR CLASSIFICATION IN MICE, USING PIEZOELECTRIC PRESSURE SENSORS

By

Steven Ryan Gooch

Kevin Donohue
Director of Thesis

Caicheng Lu
Director of Graduate Studies

May 5, 2014

ACKNOWLEDGMENTS

The following thesis benefited from the assistance of several people. First, my research advisor, Kevin Donohue, was consistently and massively helpful in all facets of not just my thesis work itself, but in the research process, teaching pursuits, and underpinning theory. Second, Bruce O'Hara was instrumental in providing an understanding of the behavioral and biological information needed in this work. Next, the lab of Sridhar Sunderam provided an immense and timely collection of data which this thesis work tested upon, in addition to key insights into the data and theory as well. In regards to the hefty work of behavior scoring and data collection used in this study, I would like to specially thank Farid Yaghouby and Chris Schildt. Finally, I would like to thank my Thesis Committee: Kevin Donohue, Bruce O'Hara, and Laurence Hassebrook. Each provided useful comments and questions throughout the pursuit of my master's degree and were instrumental in improving the finished thesis.

In addition to the technical assistance above, I would like to thank my parents for their consistent support at all points in my educational career. Karlin Levine-Smith provided non-technical assistance in the writing process and as a sounding board for research ideas. Finally, I wish to thank the friends and family who persevered with me throughout this process and never yielded their support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS III

LIST OF TABLES VI

LIST OF FIGURES VII

CHAPTER 1 – INTRODUCTION 1

1.1 THE STUDY OF MOUSE BEHAVIOR 1
1.2 INTRODUCTION TO THE SYSTEM 2
1.3 LITERATURE REVIEW 3

CHAPTER 2 - METHODS 12

2.1 INTRODUCTION 12
2.2 CAGE SYSTEM 12
2.3 EEG/EMG ACQUISITION 12
2.4 VIDEO OBSERVATION OF ACTIVE BEHAVIORS 13
2.5 THE CLASSIFIER SYSTEM 14
2.5.1 Classifier Program Flow 14
2.5.2 Linear Discriminant Analysis Metric 15
2.5.3 Mahalanobis Distance Metric 16
2.5.4 Classifier Sensitivity Rate Calculation 17
2.6 MICE STRAINS 18
2.7 BEHAVIOR DESCRIPTIONS AND SIGNAL PROCESSING 18
2.7.1 Data Processing 19
2.7.2 Non-Rapid-Eye Movement (NREM) Sleep 20
2.7.3 Rapid Eye Movement (REM) Sleep 21
2.7.4 Quiet Rest (REST) Behavior 22
2.7.5 Rearing (REAR) Behavior 23
2.7.6 Locomotion Behavior 24
2.7.7 Eating (EAT) and Drinking (DRINK) Behavior 25
2.7.8 Grooming (GROOM) Behavior 27

CHAPTER 3 – ALGORITHMS AND FEATURE SELECTION 28

3.1 INTRODUCTION 28
3.2 POWER SPECTRAL DENSITY 29
3.3 AUTOCORRELATION 30
3.4 SIGNAL ENVELOPE 31
3.5 RAW SIGNAL FEATURES 33
3.6 POWER SPECTRUM FEATURES 36
3.7 AUTOCORRELATION FEATURES 39
3.8 SIGNAL ENVELOPE FEATURES 42

CHAPTER 4 – CLASSIFIER DESIGN AND PERFORMANCE 44

4.1 INTRODUCTION 44
4.2 SYSTEM OVERVIEW 44
4.3 METHODS AND TESTING 46
4.4 SLEEP/WAKE CLASSIFIER 50
4.5 NREM/REM CLASSIFIER 56
4.6 REST/ACTIVE WAKE CLASSIFIER 62

4.7 LOCOMOTION/MEDIUM ACTIVE WAKE CLASSIFIER 65
4.8 CONCLUSION 67

CHAPTER 5 - CLASSIFIER PERFORMANCE IN THE PRESENCE OF NOISE, AND COMPARED TO OTHER CLASSIFIERS 70

5.1 INTRODUCTION 70
5.2 NOISE TESTING 71
5.2.1 Additive Gaussian White Noise *71*
5.2.2 Noise at Specified Frequency *73*
5.3 ALTERNATIVE CLASSIFIER CONFIGURATIONS 75
5.3.1 Three Class System *75*

CHAPTER 6 – CONCLUSIONS AND FUTURE WORK 79

6.1 CONCLUSIONS AND FUTURE WORK 79

REFERENCES 82

VITA 86

LIST OF TABLES

Table 3.1 Feature superset list 28
Table 4.1 SLEEP/WAKE classifier feature set..... 51
Table 4.2 SLEEP/WAKE classification results. 53
Table 4.3 SLEEP/WAKE system accuracy values for each subject, classified by LDA. 54
Table 4.4 SLEEP sensitivity values for each subject, classified by LDA. 55
Table 4.5 WAKE sensitivity values for each subject, classified by LDA. 56
Table 4.6 Feature set used to classifiy REM and NREM substates of sleep, with Fisher's Linear Discriminant values. 57
Table 4.7 NREM/REM classification results..... 59
Table 4.8 Mean NREM/REM classification accuracy for each subject. 60
Table 4.9 NREM sensitivity values for each subject. 61
Table 4.10 REM sensitivity values for each subject. 62
Table 4.11 Low activity WAKE versus active WAKE classifier features. 63
Table 4.12 REST/Active WAKE classification results. 64
Table 4.13 Linear discriminants for the features in the high active WAKE and medium active WAKE classifiers. 65
Table 4.14 Classifier results for the high active (LOCOMOTION) and medium active (GROOM, EAT, DRINK, and GROOM) substates. 67
Table 5.1 Classifier success rates in the presence of white noise, as categorized by signal to noise ratio (SNR). 72
Table 5.2 Fisher's Linear Discriminant values in the presence of Gaussian white noise. 73
Table 5.3 Classifier success rates in the presence of an 8 Hz noise signal at varying SNR values. 74
Table 5.4 Fisher's Linear Discriminant values in the presence of 8 Hz white noise, with SNR of -3.92. 75
Table 5.5 Feature set used in the NREM-REM-WAKE classifier. 77
Table 5.6 Confusion matrix demonstrating sensitivity rates for the three behaviors. 77

LIST OF FIGURES

Figure 1.1 Decision tree classification scheme. 3
Figure 2.1 Characteristic NREM sleep signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 21
Figure 2.2 Representative REM sleep signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 22
Figure 2.3 Characteristic signal representing the quiet wakeful rest state of the mouse. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 23
Figure 2.4 Characteristic REAR mouse behavior signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 24
Figure 2.5 Characteristic LOCOMOTION behavior signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 25
Figure 2.6 Characteristic EAT behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 26
Figure 2.7 Characteristic DRINK behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 26
Figure 2.8 Characteristic GROOM behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz. 27
Figure 3.1 Power spectrums of (a) SLEEP and (b) WAKE piezo signals. 30
Figure 3.2 Autocorrelation plots of sleeping (a) and waking (b) piezo signals for mice. 31
Figure 3.3 Plot showing 1 second of a LOCOMOTION piezo signal with its computed envelope overlaying the signal. The envelope is always greater than or equal to the original signal (solid line). 32
Figure 3.4 Examples of normalized pressure signals. (a) Mouse is in REM sleep. (b) Mouse is moving across cage floor. The peakedness value in the REM segment is 177.97, while the value in LOCOMOTION is 148.56. 37
Figure 3.5 Power spectra of (a) LOCOMOTION behavior piezo signal, and (b) REM sleep piezo signal, right. 40
Figure 3.6 Comparison of autocorrelations of differing behavior signals. (a) NREM Sleep signal. (b) Rearing piezo signal.. 42
Figure 3.7 Signal envelopes. (a) Grooming behavior signal envelope. (b) NREM sleep signal envelope. 43
Figure 4.1 Overall decision tree system. Vertical columns represent stages in the classification system. 45
Figure 4.2 Alternative representation of the decision tree classifier system. 46
Figure 4.3 Overall decision tree classifier system with percentages of classification successes, including propagated errors from previous steps. 68

Chapter 1 – Introduction

1.1 The Study of Mouse Behavior

Mice share most genes and gene functions with humans and other mammals, making them the preferred, low-cost animal to study pharmaceuticals, diseases, behaviors, and activities in relation to humans. It has been estimated that 98% of the mouse genome is homologous to the human genome [1], so the effects of tests are likely to be similar to human responses in many tests. Mice are also small, relatively inexpensive, and easy to maintain, as well, which allows for lower cost and higher throughput studies than the use of larger animals allows. It is for these reasons that the use of mice for laboratory testing is so widespread, a fact that is unlikely to change as research progresses in time.

The most popular technologies used to study sleep and wake in mammals are Electroencephalograms (EEGs), Electromyograms (EMGs), and Electrooculography (EOG) tests. In mice, EEG combined with EMG is the norm [2], but as these methods involve surgical implantation of electrodes, they are high cost and thus impractical for large-scale studies. To maintain the high throughput, large-scale studies often needed for pharmaceutical, genomic, and behavioral research, non-invasive methods need to be examined.

This study was focused on a non-invasive method to classify both active and resting mouse behaviors. The behaviors of interest were Rapid Eye Movement (REM) sleep, Non Rapid Eye Movement (NREM) sleep, quiet wakefulness (REST), active behaviors (locomotion, grooming, rearing), and feeding behaviors (eating, drinking). This

work develops and evaluates a novel, hierarchical minimum distance classifier method for distinguishing these behaviors based on recorded mouse pressure signals.

1.2 Introduction to the System

The system used to capture the movements of these mice has been well-defined elsewhere [3][4], but a brief summary is worth developing here. The mice are placed in a cage, with some bedding and ready access to water and food, and atop a polyvinylidene fluoride (PVDF) sensor that covers the entirety of the base of the cage. This sensor pad measures the variations in pressure caused by the mouse in the cage, and is sensitive enough to detect even the slightest perturbations due to respiratory movements. The signal then passes to an input differential amplifier and a low pass filter, which effectively provides band pass filtering. This filters the DC and low-frequency interference, preventing amplifier saturation. The signal output amplitude of the amplifier is passed to a National Instruments DAQ card (NI-DAQ X6341), which digitizes the signal at a sampling rate of 128 Hz and quantizes the signal at 16 bits. Data processing, classification, and data analysis were implanted using MATLAB R2012a (Mathworks, Natick, MA).

This thesis will develop and evaluate an automated non-invasive method for behavior classification in mice. The classification utilized a hierarchical approach involving cascaded binary classifiers to differentiate mouse behaviors, and was compared to a multi-state approach for comparison and validation. The features used in the classification were extracted using algorithms including the Power Density Spectrum [5] and Autocorrelation [5]. The quality and performance of the features will be discussed, with respect to the implemented linear discriminant analysis (LDA) classifier system. The

LDA classifier offers one main advantage with respect to the minimum distance metric used in the multi-state classifiers presented later; the LDA offers the movement of the decision threshold, while the minimum distance assumes both classes occur with equal frequency, which is not true in many of the cases below.

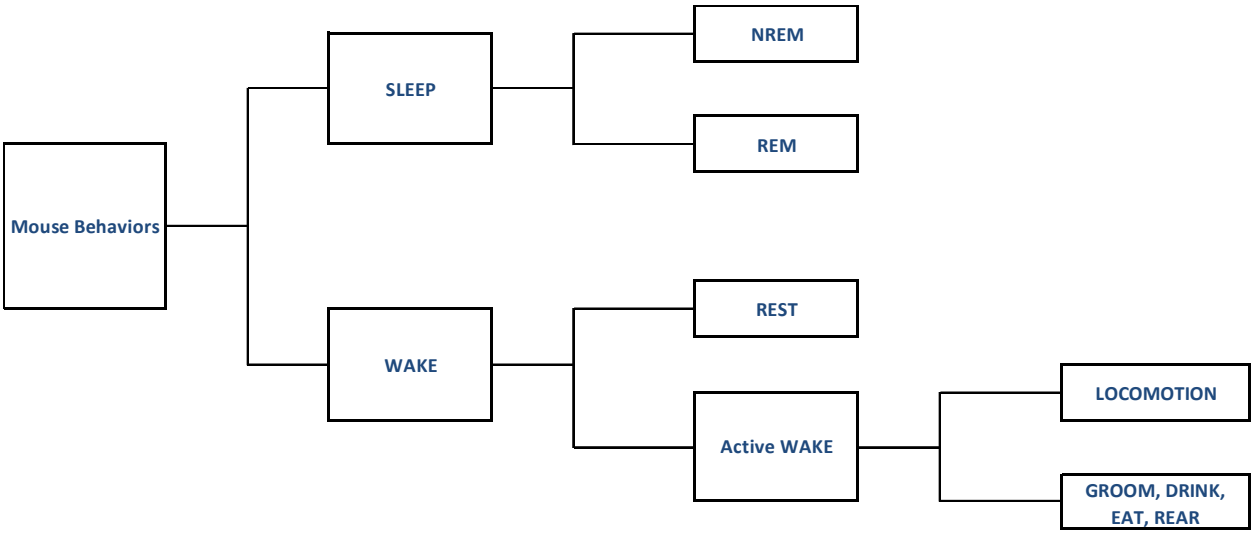


Figure 1.1 Decision tree classification scheme.

1.3 Literature Review

Alternatives to the expensive, automatic classification of mice behaviors have been of interest to researchers for decades. The first team to implement a PVDF pressure sensor, as presented in this thesis, was Megens, et al [6], which was focused on the effects of various drugs on rat behavior. The sensors used included two parallel sensors, and it was shown that piezo pads could detect changes in activity, though no behaviors were discerned in particular. In addition, respiratory movements were filtered from the signal, which are of major interest to discerning the stages of sleep, as discussed later here. In contrast, Flores, et al [7] reported a 95% success rate in classifying sleep from

wakefulness, implementing a full PVDF sensor attached atop the floor of a cage, and using a neural network classifier to discern the behaviors. The success rate is impressive, but the study had a limited data set, making its contribution to the training of the neural network and the intuition behind the successful classification even less clear. The system presented in the present thesis would counteract this issue by implementing an LDA classifier, which has been shown to successfully classify behaviors in mice using PVDF sensing. Also, as such systems focus on discerning only the sleep behaviors from waking behaviors, their scope is limited with respect to the goals of the present text.

In 2008, Donohue, et al [3] presented a noninvasive sleep-wake classifier for mice that utilized a PVDF sensor and an LDA classifier and achieved a classification rate of 94%. The system used 5 features, which were chosen to exploit the periodic signal traits inherent in sleeping mice. To yield the highest classification rates, the behavior segments were stored in 8-second segments and amplitude compression and tapering were applied to the signal. Validation was provided by human observer scoring, where observers recorded whether a mouse was asleep or awake. The success rate for this work is high, but the study was limited by its small data set, using only 4 mice and cataloging only 28.5 hours of data. However, the method did show the feasibility of using a piezoelectric pressure sensor to classify sleep and wake behaviors in mice and forms the basis for the analysis in this thesis. A second paper by Donohue, et al [8] examined the use of another classification method to differentiate sleep and wake states. A Support Vector Machine (SVM) was designed using the top 5 features as produced by a genetic algorithm search of 26 potential features. Again, human observation and visual scoring were used for validation, and the data set is generated from observation of 4 mice. In the end, the SVM

was responsible for a 95% classification agreement with human observers. This validated the use of LDA in classification of mouse behaviors, as the SVM produced almost identical results with the LDA method. Again, however, the study suffered from a small sample size, and covered only 14.5 hours total of data from the 4 mice.

Other methods have been examined for non-invasive automated behavior detection in mice, with primary focus on the classification of sleep from wake, and restful activity from motor activity. Kjellstrand, et al [9], used a Doppler Radar system to measure activity. The system converted the mouse-initiated Doppler shifts to an analog signal, and when a detected shift was greater than a certain preset level, a count was registered. For some high energy activities, such as walking, rearing, and running, the system recorded higher signal levels accurately. However, this was in large part due to the shaking of the cage itself. High energy activities like grooming, thus, were often missed by the system, when the mouse was stationary, and the cage was motionless. The study also made no attempts with discerning the restful activities of sleep and quiet rest. Marsden, et al [10] also attempted to use a Doppler system to measure locomotor activity in rats. Their system directed the microwave radiation from above, and two frequency bands were used in order to differentiate low speed activity from high speed activity. The study concluded that Doppler radar could be useful in detection of active behaviors. However, there was no direct observational component in the testing to verify that the system was in fact recording what was expected. The authors claim that the diurnal nature of the 24 hour signal verified its accuracy, but offer no quantitative results. While, unlike in the previous study, the lower frequency band detected lower speed motions and thus picked up behaviors like grooming, no effort was made to discern low energy

activities like sleep again. Zeng, et al [11], addressed these low energy behaviors, using a Doppler radar unit and support vector machines (SVM) to classify wakefulness from non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. In order to verify their successful classification of these sub-sleep behaviors, two trained observers scored EEG and EMG data for the three behaviors. The team reported 83% sensitivity agreement classifying wakefulness, 92% sensitivity agreement for NREM sleep, and 62% sensitivity agreement for REM sleep. It was hoped that respiratory rates, gross body movements, and heart rates would be usable for classification, but the study could not reliably detect heart rates. Neither is it clear that the radar detected only the respiration of the rat itself, as EEG/EMG cables were in the cage and affixed to the rat as well. The oscillation of the cables could, in certain postures, be what was detected by the radar and not the rat's breathing instead. The study incorporated thermal analysis as well, but it was found that body temperature between NREM and REM was indistinguishable, so the 23 features used in the SVM came from using only respiratory rates and body movements, similar to the system presented in this thesis. As evidenced by the lower rate of success, fewer features were found that adequately described REM. Finally, the computational expense of SVMs is much greater than that of the minimum distance classifiers described below.

Other non-invasive methods have been tested for reliability in automated behavioral classification as well. Clarke, et al [12], introduced an infrared array system to detect locomotor activity in small animals. Essentially, the system consisted of infrared beams placed 3.0 cm above the cage floor, directed horizontally through the cage. As the animal moves, certain beams would be selectively occluded, which in turn could be converted to a voltage and counted to form an activity signal. The system worked well to

detect gross motion, but the apparatus was not designed with any consideration of discerning stationary activity from stationary inactivity, limiting its usefulness. This is an issue all infrared systems face, limiting their usefulness as a sole detector for any type of animal activity.

Hilakivi, et al [13], studied the static charge sensitive bed (SCSB) system and its usefulness for detecting sleep and wake behavior in newborn rats. SCSBs are similar to the Piezo sensors presented in this thesis, in that the subject is in contact with and atop the sensor pad itself. The signal is obtained, however, by measuring the redistribution of static charge due to distorted conducting layers in the sensor caused by weight shifting by the subject itself. The signal was passed through a differential preamplifier, then filtered into two output signals, including a signal for respiratory motion and a signal for total movement. EMG and EEG recordings were made, in addition to observational scoring of the rats, to test the SCSB system's success in classification. The paper reported moderate success by the SCSBs in classifying active sleep, quiet sleep, and waking, but the classification itself was limited in that humans were still required to score the sensor readings. While the method was non-invasive, it was still necessary for observers to be trained and to take the time to score the SCSB signals. Additionally, the agreement between scorers was reported lower than is typical for EEG/EMG scoring, at 77% agreement. Also, the scores were often in disagreement with those scored from observation of the rats and the EEG/EMG scores, limiting the effectiveness of the method. Also, no effort was made to reduce any potential crosstalk component whereby reverberations in the cage due to external or internal system motion could translate to the SCSB sensors themselves.

Päivärinta, et al [14], sought to expand the usage of SCSBs to active behavior detection and implemented a system to detect fighting behaviors between male mice. The systems were mounted in an isolated manner from external vibrations, to reduce noise from external sources. Glass was placed atop each SCSB pad to spread out evenly pressure, to make a more uniform signal for analysis. The paper defined various behaviors as fighting and used minimum intervals to better ascertain the data desired. Signals were separated using these latencies. The results were tested against observationally scored video recordings for accuracy. When fighting events were observed, the SCSB correlated highly with the observational scoring data. However, it was found that roughly half of the fighting events were missed due to the high level of thresholding, leading to misclassification and an error rate of 50% for fighting behaviors. The system was limited by its latency intervals, and choosing the balance between latency intervals and successful classification of behaviors may limit the overall value of the system.

Another system that has been studied is biotelemetry, researched by Gegout-Pottie, et al., [15] for the use of detecting continuous locomotor activity in rodents. In their system, a transmitter was placed in the abdominal cavity of a rodent that sends data to a receiver placed below the cage. The method reduces the need for cables attached to the rodent, as in typical EEG/EMG studies, thus reducing the impact on behavior that such cables have on subjects. The experiment relied on the understood nature of temperature variations with respect to activity in rodents and its ability to predict the mobility of the animal. As the study admits, this method assumes a high level of correlation between body temperature change and LOCOMOTION, but there are a

number of factors that could affect this, and as such, there is no linear function that connects the two. Tang et al [16], in a study that was focused on comparing sleep and locomotive activity between different strains of mice, also implemented a telemetric system of the nature described above. It was reported that the system had the capacity to record heart rate, body temperature, and brain temperature, but due to size constraints, only EEG signals and gross whole-body activity signals were recorded. This activity signal was in the form of TTL pulses that were recorded by the magnetic receiver plates beneath the cage apparatus. The system also featured video and infrared photobeam analysis to compare results obtained by telemetry. It was found that the TTL pulses appeared only in the waking segments, consistent with the idea that mice move very little when asleep. The results were promising in that EEG signals and whole-body movements could be recorded with accuracy and without obstructive cables, but the system does still require the extensive and invasive surgery this thesis hopes to eradicate. Also, as the gross whole-body movement detection needs the magnet to be moving spatially over the receiver, active behaviors like GROOM and feeding are lost 10 the system since the mouse remains relatively stationary for the durations of these behaviors.

Working on the magnet method, Storch, et al [17] sought to classify sleep wake behavior in mice with a subcutaneously implanted magnet, placed in the neck muscles. There was no telemetric reporting of results, and surgery would be less intensive for such a method. The motion of the magnet in space was tracked by a custom magneto-sensitive device with magnetic field sensors placed evenly in a grid beneath the cage. When placed in the neck muscle, behaviors such as grooming were discernible due to the periodic nature of the motion of the head in that behavior with respect to the cage floor sensors.

However, the transition to sleep was not easily recorded by this method, as in the state of quiet restfulness, the head moves nearly as little as during NREM sleep. Thus, the group had to use a defined inactivity limit for sleep onset, which could skew results. The paper also presents only that the EEG scores and magnet discerned similar amounts of sleep, but not that the same sleep and wake epochs were discerned in each case. These factors, along with the need for a surgical implantation, limit the effectiveness of this method for classification.

Video analysis has been an area of interest in behavior research, as seen in a few of the implementations above. Pack et al [18], showed that high-throughput sleep scoring could be achieved using video monitoring, infrared sensing, and an object recognition algorithm that expanded upon the ideas that sustained levels of inactivity, at a certain point, correlated highly with sleep. Video recordings generally had been more useful in classification of active behaviors, as moving mice were easy to detect with such analysis. Fisher et al [19] sought to challenge this notion, however, and studied the use of video tracking systems for determining sleep versus wake behavior. Their system utilized a wide angle lens and an off-the-shelf video tracking software installation. As a means of comparison, EEG/EMG signals were obtained using telemetric transmission. As in the Pack et al. study above, sleep was defined using immobility for a certain duration of time, but unlike that study, an additional feature involving the amount of the subject required to be stationary for sleep to occur. This way, sleep could be discerned from other stationary behaviors, such as sustained EAT or DRINK. Using a method for establishing agreement between EEG/EMG recordings and Video tracking results [20], 95% agreement for the detection of sleep and wake was recorded. The system does provide a high throughput,

accurate system, but as the detection of REM and NREM sleep is not described, the study comes up short in comparison with the goals of this thesis.

In conclusion, multiple alternatives have been studied to automatically detect behaviors in mice, but as seen above, no method has been attempted that truly offers an inexpensive, high throughput classification scheme that studies both inactive and active behaviors, and discerns NREM from REM sleep at once. The goal of this thesis is to study a system that could satisfy those goals.

Chapter 2 - Methods

2.1 Introduction

The mouse data acquisition system used in this work was based on the efforts of previous works [1][2], with a few alterations and upgrades. This chapter will discuss the acquisition methods used to acquire the data used in testing and training the developed classifier.

2.2 Cage System

The cage used is a modification on the Mouse Sleep-Wake Tracking System (Signal Solutions, Lexington, KY) and is part of a four-cage module for the testing of four mice at once. Cages are (6x6 inches), with bedding on the cage floor for the comfort of the mice, as well as readily accessible food and water.

A PVDF sensor SP 77 (Signal Solutions, Lexington, KY) rests on the cage floor and is used to convert pressure changes resulting from mice movements, and connected to a Sensor Amplifier PZA 100 (Signal Solutions, Lexington, KY) for reading, amplifying, and filtering the signal. The signal is then acquired, quantized and digitized by an NI-DAQX-6341 (National Instruments, Austin, TX). A software package, Sirenia Sleep (Pinnacle Technology, Inc, Lawrence, KS) is used to then store the piezo voltage signal for further manipulation. Only a single channel was used in the recording and the data acquisition was modified from the typical single ended input to a differential ended input.

2.3 EEG/EMG Acquisition

Electrodes were surgically implanted in anesthetized mice, with a cable connected for recording EEG and EMG signals. The EEG signals and EMG signal were acquired using the same program as the above piezo film recordings, in Sirenia Sleep (Pinnacle).

The signals were acquired concurrently with the piezo recordings and shared the same clock. EEG/EMG signals were scored for three behavior states, including WAKE, NREM sleep, and REM sleep, by one of two trained observers, whose inter-scoring agreement rate was 90-95%. These signals were used to identify NREM and REM 4-second epochs in concurrence with the piezo signal, in order to identify features that would separate the states for use in the proposed classifier.

2.4 Video Observation of Active Behaviors

Video was recorded by Sirenia Sleep, and synchronized with the Sirenia clock. Human observers scored the data by inputting into an Excel (Microsoft, Redmond, WA) spreadsheet the times and types of behaviors, sorted into one of 8 possible behaviors: Sleep, Quiet Rest, Rearing, Locomotion, Drinking, Eating, Grooming, and Indeterminate. The observational data was of particular interest to the WAKE behaviors in the classifier design. The Excel behavior data was imported to Matlab (Mathworks, Natick, MA) using a custom program, and behavior times and lengths were converted to a continuous signal simulating unique voltages for each behavior type. The piezo data was exported from the Sirenia Sleep program in the European Data Format (EDF), which was in turn imported to Matlab and lined up with the simulated voltage signal. A program then parsed through the simulated voltage signal, and extracted corresponding piezo data segments and put them into their own behavioral matrices. A guardband of 8 seconds was implemented in order to define behavior segments of sufficient length that were exactly one type of behavior. Piezo data segments that satisfied this requirement were kept and made up the behavior matrix that was studied for the extraction of features for the waking behaviors.

2.5 The Classifier System

2.5.1 Classifier Program Flow

Overall, as seen in Fig. 1, the behavior state and substate classifier presented in this work follows a decision tree pattern. Each "branch" in the tree is the result of a binary classification step, where an LDA classification was implemented. After the piezo data was acquired and converted to a usable format in Matlab, and placed into matrices based on the behavior that generated the signal, a script was run to compute the features relevant for classifying behaviors at each level in the hierarchical scheme presented in Fig. 1. Signal Segments labeled by human EEG scoring and visual observation were organized column-wise in behavior matrices.

These segments were individually passed to a function which calculated each of the features desired, afterwards returning the features to the script in the form of a feature vector. Column feature vectors were accumulated in behavior feature matrices and stored for use in the designing and testing of the classifier.

To generate the training and testing matrices, feature vectors were randomly sampled from each behavior feature matrix, and used to populate the respective behavior training matrices. In each classifier implemented in the decision tree system, at least 100 unique feature vectors were used to populate the training feature matrices respectively. The test set populated by at least 100 feature vectors from each of the remaining behavior feature matrices. These were selected such that no training feature vector could be used as a test vector, and likewise no testing vector could be used for training the classifier.

Mean vectors were calculated for the training and testing feature matrices, and used as class means. The inverse covariance matrices were calculated as well, which were used in the LDA classifier.

2.5.2 Linear Discriminant Analysis Metric

Unless otherwise specified, all classifiers in this paper utilize LDA to classify the behavior pressure signals. LDA was chosen for several reasons. First, all classifiers in the hierarchical system are binary in nature, and second, the LDA classifier allows for the manipulation of the decision threshold, which is useful in controlling error type. This is especially useful in the cases where one state is overwhelmingly more common than another; a minimum distance classifier, with its threshold at zero, would be less optimal as there is an inherent assumption that both classes occur with equivalent frequency.

Linear discriminant analysis was described first by Fisher [4]. Sets of features $\mathbf{f_1}$ and $\mathbf{f_2}$ are chosen from classes $\mathbf{y_1}$ and $\mathbf{y_2}$ and computed for observations in each segment. Training sets, $\mathbf{x_1}$ and $\mathbf{x_2}$, are populated using segments from the feature sets. Mean feature vectors, $\bar{\mathbf{x}}_\mathbf{1}$ and $\bar{\mathbf{x}}_\mathbf{2}$, are computed by averaging the values for each feature across all segments in the training sets. The goal of this process is to obtain a set of linear weights of size 1xN, where N is the number of features in the feature vectors, which can be used to weight the features in a test vector to compute a value which, if greater than a certain threshold, corresponds to one class, and if less, then the other.

These weights are obtained by computing the covariance matrix of the mean vectors, $= \left[\bar{\boldsymbol{x}}_\mathbf{1}^T\, \bar{\boldsymbol{x}}_\mathbf{2}^T\right]^T$. The covariance matrix C is given by

$$\mathrm{C}[\mathbf{X}] = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] \tag{2.1}$$

where $E[\mathbf{X}]$ is the expected value of the mean feature matrix $\mathbf{X}$. The result is an NxN matrix, sometimes referred to as the pooled covariance matrix, since it is the product of the feature matrices from both classes. Once this is obtained, the inverse of the covariance matrix is computed and multiplied by the difference in the mean feature vectors, to generate the linear weights $\mathbf{w}$.

$$\mathbf{w} = \mathbf{C}\, x\, (\bar{\mathbf{x}}_\mathbf{1} - \bar{\mathbf{x}}_\mathbf{2})^T \tag{2.2}$$

The LDA classifier operates on a few key assumptions. First, it is assumed that the conditional probability density functions, $p(\mathbf{x}|\mathbf{y_1})$ and $p(\mathbf{x}|\mathbf{y_2})$ are normally distributed, and that the covariance matrices of each class are equivalent, or, $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$. This assumption of homoscedasticity allows the decision metric to simplify to

$$\mathbf{w} \cdot \mathbf{x} \lessgtr \mathbf{T} \tag{2.3}$$

where **T** is the threshold and **x** is the test feature vector.

2.5.3 Mahalanobis Distance Metric

The Mahalanobis distance (MD) is similar to the Euclidean distance (ED) metric, but has a few advantages that make it advantageous for the present work. The Mahalanobis distance [6] is given by

$$MD = \sqrt{(\mathbf{x_i} - \bar{\mathbf{x}})\Sigma_x(\mathbf{x_i} - \bar{\mathbf{x}})^T} \tag{2.4}$$

In the above relation, $\mathbf{x_i}$ is the feature vector of the signal to be classified, $\bar{\mathbf{x}}$ is the mean feature vector template for a given class, and Σ_x is the covariance matrix generated from the training feature vectors for the particular class X.

One particularly useful characteristic of the MD is found in the use of the covariance matrix in determining the distance to the mean. This effectively weights features and reduces the effects of correlation between features in finding the minimum MD among a set of classes. It is easiest to visualize the effect this has on distance computation by imagining the case of a two dimensional feature space.

In calculating the ED, every point on the edge of a circle centered at the class mean is equidistant from said mean location. Meanwhile, in calculating the MD, the

feature covariance weighting makes it such that every point on the edge of some ellipse, with its centroid positioned at the class mean, is equidistant from the class mean. However, if the cross correlation between every set of two features is zero and class covariances are equal, then the Mahalanobis distance reduces to the Euclidean distance. This occurs when every off-diagonal entry in the covariance matrix Σ_x is zero. Discussion follows concerning the success rates and how these were calculated in the next subsection.

2.5.4 Classifier Sensitivity Rate Calculation

Classification results in this work are generally reported for each class individually and the metric used to convey classification success is the sensitivity. The sensitivity rates [5] can be interpreted as the average value of the number of feature vectors classified appropriately in their class, divided by the total number of feature vectors in the test set for a particular behavior.

$$S = \frac{TP}{TP + FN} * 100\% \tag{2.5}$$

In the above relation, S is the sensitivity, TP is the true positives for some behavior, and FN is the false negative classified segments. A bootstrapping process was used to design and test many classifiers. The rates reported in this work are the average sensitivity rates resulting from 200 Monte Carlo runs of each classifier, in which each run was subject to randomly sampled training and testing behavior feature vectors. This repeated testing lends more variability to the sensitivity rates as actual rates of classification, with some of the intrinsic variability accounted for. So, the sensitivity rates themselves are averages derived from this bootstrapping. As such, standard deviation values could be calculated to

find the variability in the success rates for different configurations. Generally, the classifiers presented never had a standard deviation greater than 3-5%.

In some areas, the classifier system accuracy rates are reported in addition to the sensitivity rates above. The accuracy rate offers a valuable insight into the contribution of false negatives and false positives in the error values of a given classifier. There is especially a dichotomy between the accuracy rates and sensitivity rates in classifiers where the probability of occurrence of one class is much greater than the probability of occurrence of the other. The accuracy, A, is given [5] by

$$A = \frac{TP_1 + TP_2}{TP_1 + TP_2 + FP_1 + FP_2} \tag{2.6}$$

Where TP_1 is the number of true positives for class 1, TP_2 is the number of true positives for class 2, FP_1 is the number of false positives of class 1 (ie, those segments classified as class 1 that were from class 2), and FP_2 is the number of false positives for class 2.

2.6 Mice Strains

Twenty mice were used to record the 480 hours of data used in this work. Nineteen of the mice were from the C57BL/6 strain, and one subject was from the CFW strain. For each mouse, 24 hours of data was cataloged and analyzed.

2.7 Behavior Descriptions and Signal Processing

The objective of the classifier presented is to discriminate automatically between seven different mouse behaviors; non-rapid eye movement sleep (NREM), rapid eye movement sleep (REM), quiet wake (REST), high active wake or locomotion (LOCOMOTION), grooming (GROOM), eating (EAT), drinking (DRINK), and rearing (REAR) behaviors. In this section each behavior will be described briefly and accompanied by their respective piezo readings. Also featured will be the representations

of the signals in the frequency domain, and after envelopes and autocorrelations are computed. The data processing for the signals will also be discussed below.

2.7.1 Data Processing

All signals were bandpass filtered prior to analysis or classification using a 512th order finite impulse response (FIR) filter in Matlab, with a lower cutoff frequency of 0.5 Hz, and a higher cutoff frequency at 11 Hz. The lower bound allows for the DC bias to be eliminated, while the higher cutoff was utilized to reduce higher frequency interference outside the range where motions of interest have most of their energy. Mice pressure oscillations rarely exceed 11 Hz, so any frequency component above that range could be considered noise and discarded. Also, since the PVDF sensor is sensitive to ambient RF noise at the frequency of 60 Hz, the cutoff should minimize any component in the signals due to this frequency.

Where the power spectrum density (PSD) plots are shown, the following parameters were used to find the transformation. PSDs for this work were calculated using a Hamming window twice the length of the sampling rate of 128 Hz, at 256 samples. The window overlap was 128 samples, with zero padding of 512 samples. PSDs were calculated by computing the Fourier Transform of the signal and then displaying the magnitude, plotted in dB.

Autocorrelations (AC) for this work were computed using the Matlab command xcorr, with the parameter “coeff” specified to normalize the resulting signals, forcing the zero-lag amplitude to be equal to 1. This allows for relative comparisons to be made between behaviors. The plots seen below display only the positive lags to bypass the

redundant symmetry of autocorrelation transformation, with linear trends removed from the resultant autocorrelation.

In order to compute the signal envelope, the Hilbert transform was computed and the analytic signal was obtained by using the Hilbert transform as the imaginary part, the original data as the real part. From this analytic signal, the magnitude was computed to find the envelope of the signal.

2.7.2 Non-Rapid-Eye Movement (NREM) Sleep

One of the more challenging pursuits of any non-invasive method of classification is the successful detection of NREM sleep. The current method as discussed above for reliably detecting NREM sleep is via the use of electrodes. However, by careful inspection of the piezo signal in NREM sleeping mice we can extract useful features that will discern this behavior from that of REM sleep successfully.

NREM sleep in mice is characterized by rhythmic breathing [3] and thus the signal recorded by the piezo sensor should be periodic in nature, with little variation in peak heights. When the mouse is asleep, the full body is in contact with the floor sensor, allowing for the inhalation and exhalation of the mouse to be recorded as increases and decreases in signal, respectively. In NREM sleep specifically, it is these heights and the uniformity in the signal that can be exploited in a classifier system.

Figure 2 illustrates a typical NREM signal. As the mouse inhales air, its chest cavity expands, corresponding to an increase in piezo signal amplitude. Then, the mouse exhales, leading to the decreasing signal. It is clear, too, that this cycle occurs about 3

times per second in the signal below, indicative of the breathing rate in a sleeping mouse [3]. These insights can drive the feature extraction process discussed shortly.

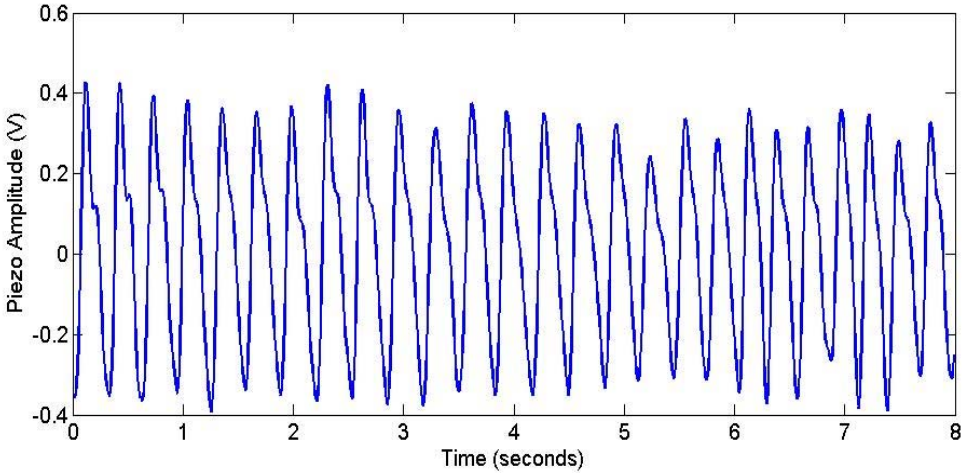


Figure 2.1 Characteristic NREM sleep signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.3 Rapid Eye Movement (REM) Sleep

In contrast with NREM sleep, rapid eye movement (REM) sleep is characterized by a different breathing frequency in many mice strains [3], as well as more erratic piezo amplitudes, as the mouse moves from deep breaths to shallow ones. In extracting useful features from REM sleep to differentiate the behavior from NREM sleep, it is useful to examine the signal's self-similarity, as well as fundamental frequency and peak height variations.

Figure 2.2 contains a characteristic piezo signal of a mouse in REM sleep. The signal looks more variable than the NREM segment, as one would expect due to the erratic nature of breathing while a mouse is in this state. Some breaths are shallower than others and some quicker than others. Also, it is clear to see the varying levels of breath intake.

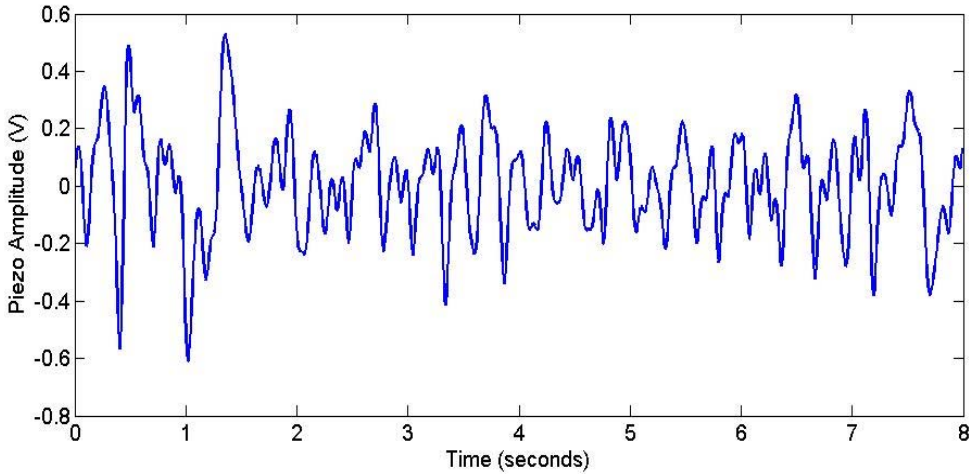


Figure 2.2 Representative REM sleep signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.4 Quiet Rest (REST) Behavior

The behavior that proves the toughest to differentiate from sleep behavior is that of quiet rest. In this state, the mouse is lying prone, with its body in contact with the piezo pad in a similar manner to that of sleep. Thus, the dominating factor driving the signal is the breathing rate. However, slight shifts in posture translate heavily to the piezo reading, as indicated by the signal spike seen in Figure 2.3. There are artifacts left of the spike that permeate the signal, leading to much more peak variation than any sleep signal as well. Also, potentially due to the posture of the resting mouse, there are small bumps in the signal on the breath intake, rising edge to each peak, a feature in the signal less often seen in the sleeping mouse.

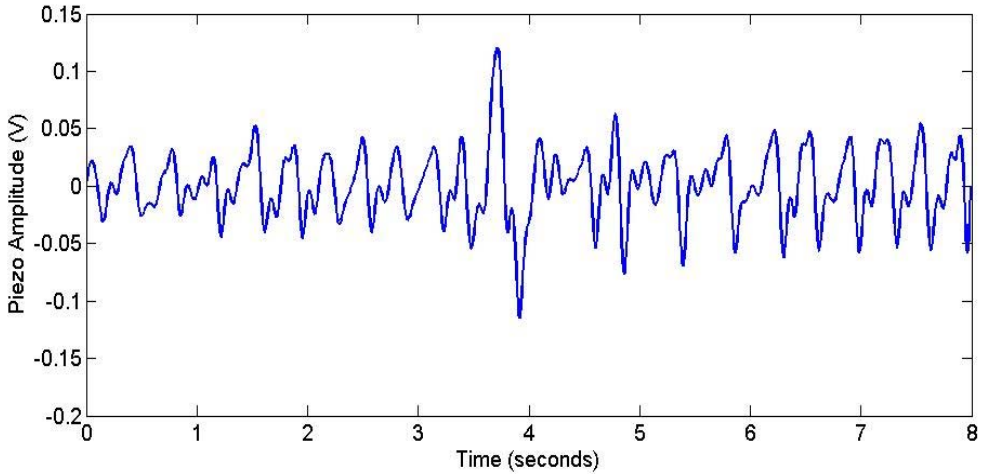


Figure 2.3 Characteristic signal representing the quiet wakeful rest state of the mouse. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.5 Rearing (REAR) Behavior

The mouse spends a significant portion of its active wake state rearing (REAR), or on its hind legs. This can be a sustained behavior or utilized in quick bouts. As such it is a difficult behavior to classify. For a system that utilizes only a piezo pressure sensor, the behavior at times can be almost indistinguishable from GROOM or LOCOMOTION behaviors, and at other times, looks similar to EAT and DRINK behaviors, where the mouse is also on its hind legs, either to reach a water spout or to use its fore limbs to manipulate the food it is EAT.

Figure 2.4 shows a characteristic REAR signal.

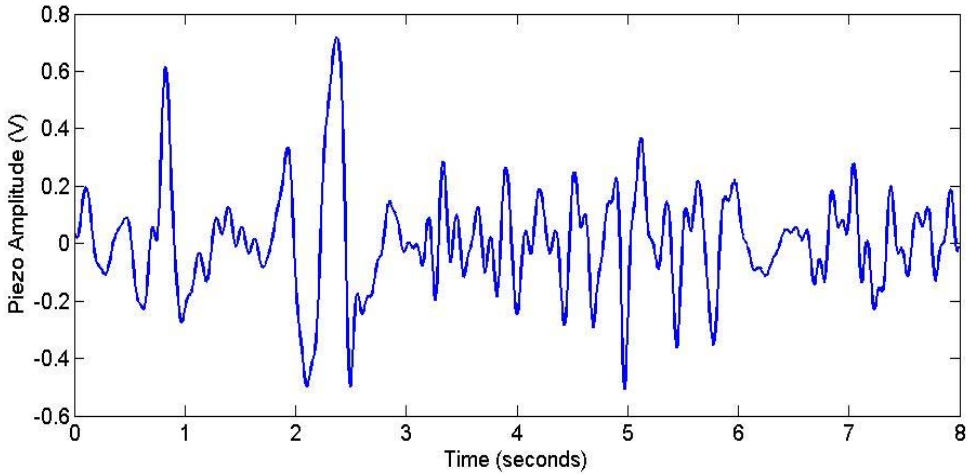


Figure 2.4 Characteristic REAR mouse behavior signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.6 Locomotion Behavior

Active LOCOMOTION refers to the behavioral pattern exhibited by mice when they are physically moving about within the cage. Generally, all four legs are in contact with the ground during this behavior as the mouse walks or runs around the floor of the cage. At times the mouse can be seen to be sniffing around and moving cautiously during this behavior as well. Some very short bouts of REAR can occur during behavior labeled as LOCOMOTION, lasting around 1 second or less in duration.

Figure 2.5 illustrates a characteristic LOCOMOTION signal for a mouse. These signals vary in amplitude and frequency unpredictably and are the least stationary of the behavior states measured. It is thus often difficult to differentiate LOCOMOTION from other active behaviors such as GROOM. Perhaps the most helpful feature is the raw signal amplitude, as the signal maximum is often higher in LOCOMOTION than any other behavior due to the mouse intensely striking the sensor while moving around.

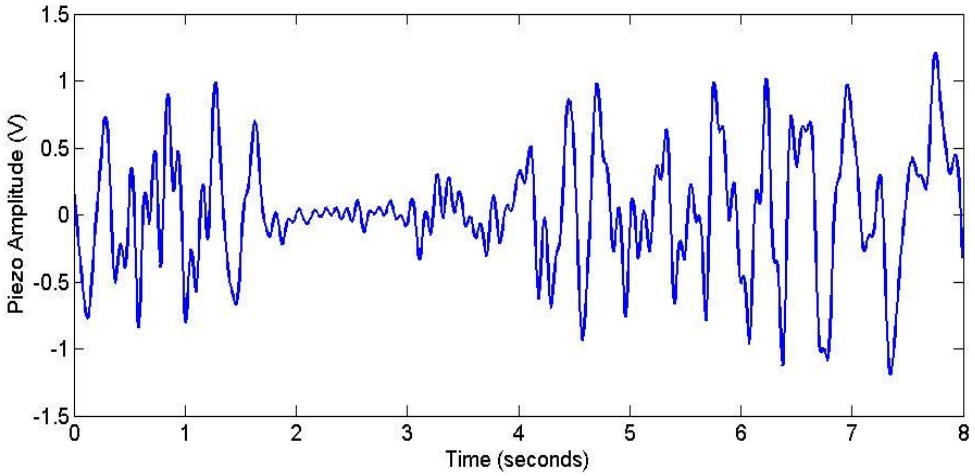


Figure 2.5 Characteristic LOCOMOTION behavior signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.7 Eating (EAT) and Drinking (DRINK) Behavior

A mouse engaged in both EAT and DRINK does so on hind legs, in this cage system. As such, it often makes sense to analyze these behaviors. However, there are subtle differences that can be detected using the floor sensor signal, and shall be described briefly below.

While engaged in EAT, it was observed that the mice reared in front of the food gate, moved their snouts and mouths into the grate to grab morsels of food, then sat down on their hind quarters to manipulate the food with their front paws and partake of the food morsels. This behavior was observed for 210 8-second epochs during one 8-hour observation study of one particular mouse, leading to a substantial dataset for testing.

This behavior prevalence was not observed, however, for DRINK behaviors. In that same 8-hour observation experiment span, for example, just 12 8-second epochs were observed where the mouse was DRINK. The DRINK signals did contain unique high frequency components, however, that could be exploited as a feature. It is surmised

that the vigorous and sustained licking required for extracting water from the spout led to these high frequency perturbations. Again, the subject needed to be on its hind legs in order to reach the water spout, leading to some signal similarity to the behaviors of EAT and REAR. Figure 2.6 shows a characteristic EAT signal and figure 2.7, a DRINK signal.

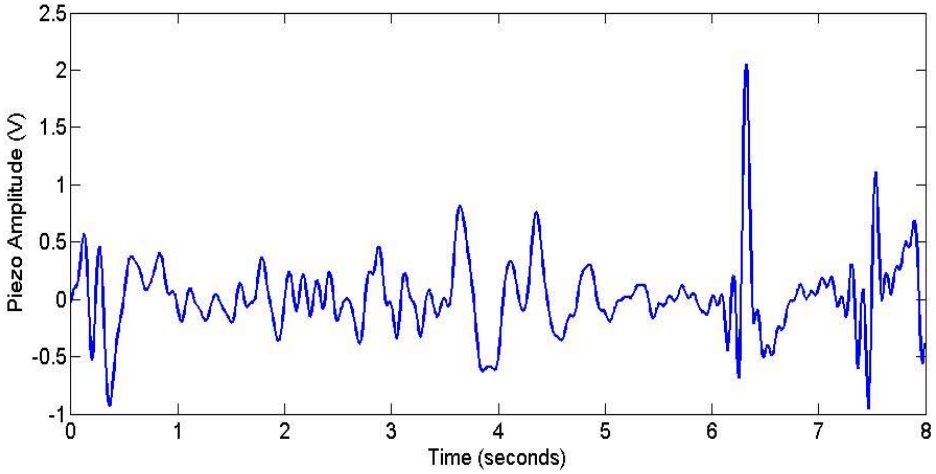


Figure 2.6 Characteristic EAT behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

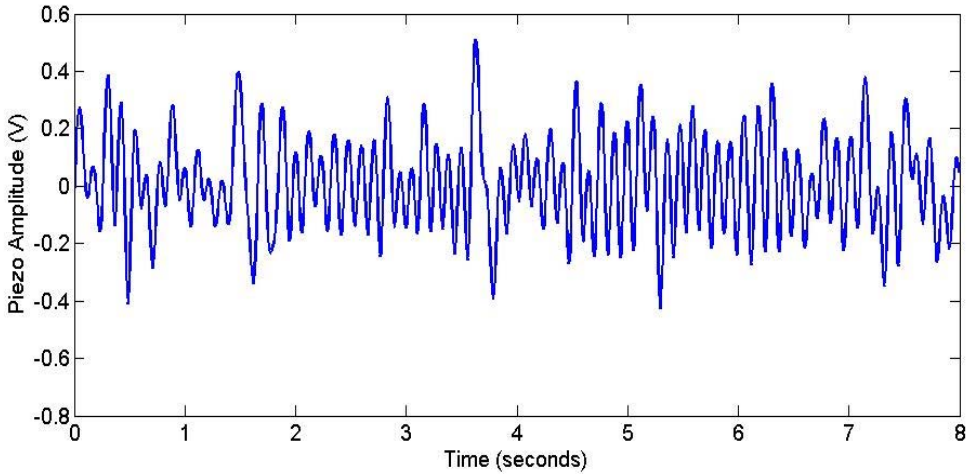


Figure 2.7 Characteristic DRINK behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

2.7.8 Grooming (GROOM) Behavior

Mice exhibit an elaborate GROOM behavior process where quite literally they clean themselves from snout to tail. Grooming signals had relatively large spikes but were often a lower amplitude behavior. The mouse does not move spatially across the cage floor so the amplitude is lower than that of locomotion, or that of rearing, where the mouse lifts fully onto its two hind legs. Instead, it remains in one position and engages in various complex repetitive motions which correspond to peaks in the signal, followed by lower signal activity. Figure 2.8 displays a characteristic GROOM signal.

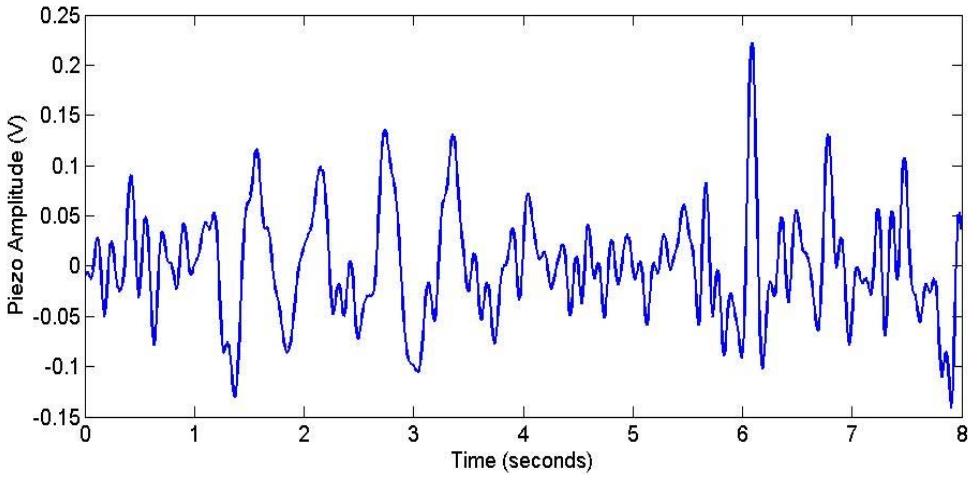


Figure 2.8 Characteristic GROOM behavior piezo signal. Bandpass filtered with cutoffs at 0.5 Hz and 11 Hz.

Chapter 3 – Algorithms and Feature Selection

3.1 Introduction

When designing a classifier, the selected features need to adequately describe the different classes from the data. It is advantageous to employ certain algorithms to evaluate the features. Along with computing features from the piezo signal, three principle algorithms were utilized, including the signal envelope, power spectral density, and autocorrelation . This chapter will discuss the algorithms used to define spaces from which features were extracted, and will also examine the value in certain algorithms for features that distinguish the various behavior states in the mice observed.

While there is an infinitude of ways to characterize signals and glean features, it was prudent in this work to limit the amount of features used to develop classifiers to a feature superset, from which smaller sets could be drawn to train the different classifiers in this work. The limited number allowed exhaustive testing of all combinations from the feature superset, yielding the best possible classifier given all combinations of features tested. In all, 20 features were used in the feature superset and follow in Table 3.1. These features will be discussed at length below.

Table 3.1 Feature superset list

Feature Number	Feature Name	Feature Number	Feature Name
1	Teager Energy (Signal)	11	Kurtosis (PSD, Frequencies 0.5 to 11 Hz)
2	Entropy (Signal)	12	THD (PSD, Frequencies 0.5 to 11 hz)
3	STD(signal)	13	Max Height (PSD)
4	Peakedness (Normalized Signal)	14	Location Max Height(PSD)
5	Standard Deviation, Peak Heights (Envelope)	15	Number Peak Heights (PSD)
6	Range, Peak Heights (Envelope)	16	Deviatedness (PSD)
7	Peakedness, Peak Heights (Envelope)	17	Standard Deviation (AC)
8	20% Ranked Value, Peak Heights (Envelope)	18	Teager Energy (AC)
9	Peakedness (PSD, Frequencies 0.5 to 11 Hz)	19	Peakedness (AC)
10	Skewness (PSD, Frequencies 0.5 to 11 Hz)	20	St. Dev. (AC) / Deviatedness Peak Heights (AC)

3.2 Power Spectral Density

The sensitivity of the PVDF sensors used in this study allows for the estimation of respiratory rates in resting mice, with the inhalation and the exhalation of breath corresponding to increases and decreases in the pressure signal read by the piezo sensor. Frequency analysis is one tool needed to examine the signals, especially with respect to sleep signals. It has been shown, in fact, that several strains in mice show differing breathing frequencies while in different stages of sleep [1]. One way to analyze signals in the frequency domain is to use the discrete Fourier transform (DFT) to find the power spectrum of a signal.

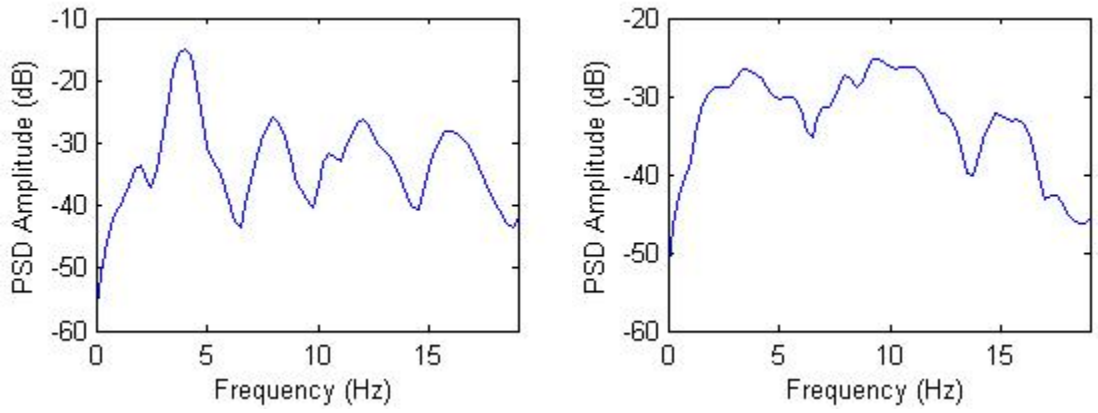
The DFT, represented below as X(f), of an input x[n], is given by

$$X(f) = \sum_{n=0}^{N-1} (w[n]x[n-N])e^{\frac{-j2\pi nf}{N}} \tag{3.1}$$

where f = 0,1,…,N

Above, the function w[n] refers to a Hamming window 2*N samples in length. The data segment windowed, x[n], was of length N samples, resulting in a 50% overlap of windows.

Characteristic power spectra obtained in this manner are plotted in Figure 3.1, representing a wake state and a sleep state. Notice that the sleep segment peaks around 3 Hz due to the subject's breathing frequency, while the wake segment has less defined peaking in the frequency domain.



(a) (b)

Figure 3.1 Power spectrums of (a) SLEEP and (b) WAKE piezo signals.

3.3 Autocorrelation

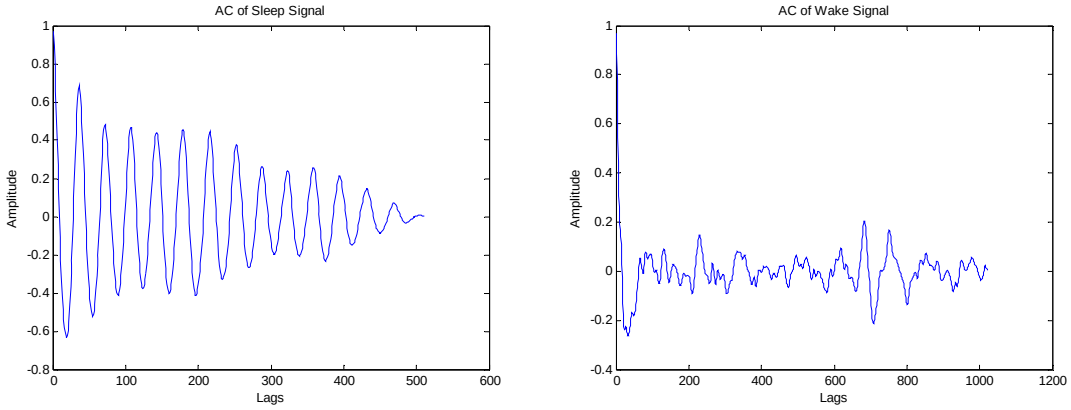
The autocorrelation (AC) is a valuable method for finding information about a signal in the time domain. For periodic signals, high values of the AC can be seen at varying lags from the original position. It can also be useful in detecting non-periodic signals, which have very low values in the AC. The extreme case of this is random noise, which has a spike at the zero lag, where every value in the noise signal line up with themselves perfectly, and zero elsewhere, as the signal never resembles itself at any other lag. This is somewhat analogous to certain active behaviors like LOCOMOTION, where changes in pressure from the feet of the mice result in a pseudo-random signal.

For a time series x(t), the autocorrelation is given by

$$r_{xx}(\lambda) = \frac{1}{s_n} \sum_{n=0}^{N-1} x[n]x[n+\lambda] \tag{3.2}$$

where $r_{xx}(\lambda)$ is the autocorrelation value at lag λ, s_n is the normalization coefficient, used to make the zero lag equal 1, and N is the number of points in the signal $x[n]$. The

autocorrelation was then normalized to allow for comparisons between signals of varying amplitudes to be made. Only positive lags are shown, due to the symmetry of the operation, and linear trends were removed prior to feature computation. Figure 3.2 illustrates the autocorrelation of two signals, a sleep signal and a waking signal.



(a) (b)

Figure 3.2 Autocorrelation plots of sleeping (a) and waking (b) piezo signals for mice. Sleep segment lengths were 4-s long at a sampling rate of 128 Hz, while WAKE segment lengths were 8-s long at the same rate, leading to the greater amount of lags in WAKE above.

3.4 Signal Envelope

In addition to the raw signal, power spectrum, and autocorrelation, the signal envelope was computed and used for feature exploitation. The analytic signal magnitude was calculated, based on the use of the Hilbert Transform, and used as the signal envelope. To obtain the Hilbert Transform, a method described elsewhere [2] was used and is described as follows. The DFT was calculated as above to find

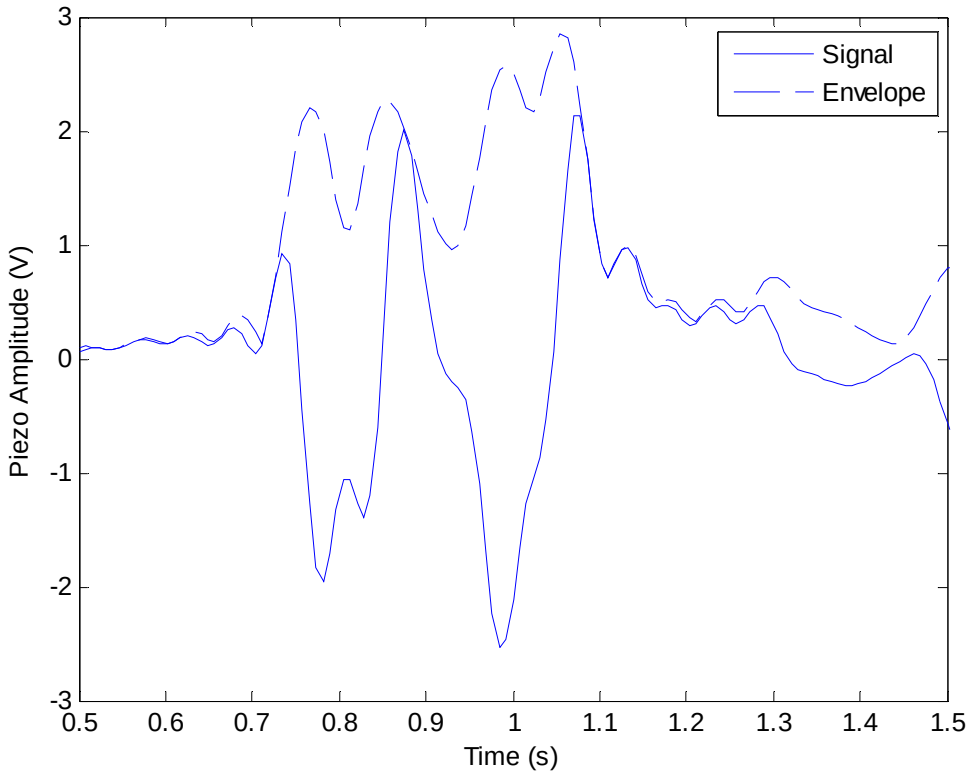


Figure 3.3 Plot showing 1 second of a LOCOMOTION piezo signal with its computed envelope overlaying the signal. The envelope is always greater than or equal to the original signal (solid line).

X(f). Then, a function Z(f) was defined such that for all negative frequency values, Z(f) was nullified, and for all positive values, Z(f) was doubled. This effectively re-distributed all the energy in the DFT to the positive frequencies. Finally, the inverse Fourier Transform [3] was computed to find the envelope function e(t). The inverse Fourier Transform is given below.

$$e(t) = \int_{-\infty}^{\infty} Z(f) e^{j2\pi f t} df \tag{3.3}$$

3.5 Raw Signal Features

Some useful features for classification can be discerned from simply evaluating statistics of the raw behavior piezo signal itself. The first feature in the feature superset described above is the Teager Energy of the raw signal. The Teager Energy operator was first detailed by Kaiser [4], inspired by the work of Teager and Teager [5][6], and defined as

$$TE = \Psi[n] = x^2[n] - x[n-1]x[n+1] \tag{3.4}$$

in discrete time. The above operation resulted in an energy signal, and by taking the absolute value of the sum of samples in that signal, the Teager Energy feature was obtained, as follows.

$$TEF = \sum_{n=0}^{N-1} |\Psi[n]| \tag{3.5}$$

Since the Teager Energy itself produces a signal, the signal average was computed as the feature used in order to have a single value for a segment's Teager Energy. It is worth noting that the input x[n] could be normalized before the energy calculation is made. Without normalizing, the operation is sensitive to signals with higher amplitudes, which would generally result in higher Teager energy values. This may adversely affect some kinds of classification operations for some subjects; in particular, system calibration is an issue. If the system is not calibrated the same amount before every test, or if different subjects are different weights, using amplitude dependent features could produce unpredictable results.

The Teager energy operator finds its value in its definition of energy. Some conventional methods find the energy in a signal by taking the signal's Fourier

Transform. By taking the absolute value squared of the different frequency bands, a metric for energy at those frequency levels can be obtained. However, this results in signals having the same amplitude, but different frequencies having the same energy, which is somewhat counterintuitive. A signal at a higher frequency contains higher energy, as energy is a function of both amplitude and frequency. Thus this method can be insufficient in evaluating the energy contained in a waveform.

Using the second order differential equation for a suspended mass object, Kaiser found a relation that showed that energy indeed is a function of amplitude and frequency. The Teager energy that he and Teager developed, however, circumvents this by taking into account the frequency as well. It can be shown, for instance, that the Teager energy of a pure sinusoid reduces to the sinusoid's amplitude squared multiplied by its frequency squared, if computed over an integer multiple of the sinusoid's period. For signals of arbitrary generation, as in the case of the mouse piezo signals, the more general form of the equation must be implemented.

This frequency dependence is useful in classifying the mouse behavior signals since the behaviors produce higher energy waveforms, in both the time domain and the autocorrelation domain, as will be seen below.

The second feature discerned from the raw signal was the Shannon entropy of the raw signal [7]. The entropy of a signal is defined as

$$H(X) = -\sum_{i=1}^{N} p(x_i) log_b p(x_i) \tag{3.6}$$

where p(x_i) is the probability mass function of the signal value range corresponding to x_i, and with b = 2. The probability density function was calculated by using a 10-bin

histogram to sort the signal values. Then, the probability that a value would fall in a given bin was calculated by dividing the sum of the values in a given bin by the total number of values in all bins. Doing this for every bin yielded a satisfactory probability mass function. This feature was useful because periodic signals, like those associated with breathing in states of low activity, have low entropy, whereas less predictable, high activity states have higher levels of entropy.

Shannon's entropy grew out of a desire to understand and quantify information gain, which is a log weighting of the probability that a certain value occurs. Adding all probability elements by their log weights is one way to quantify information contained in a signal. In this case, information gain is greater when less likely values occur, or equivalently, when the signal is less predictable. For example, the entropy of a normalized, 3 Hz sine signal of 4-second duration when computed in this manner is 2.41 bits, while the entropy of a 4-second normalized white noise signal is 3.82 bits. The greater level of unpredictability increases the entropy of the signal. Thus, it is hypothesized that entropy will be a useful feature in classifying mice behaviors.

The third feature statistic calculated from the raw signal was simply the maximum value in the raw signal. It was surmised that this could be a useful feature in classifying lower activity, as it is the upper bound on signal amplitudes in a given analysis window and can be used to detect the presence of a high intensity pad strike. As mentioned in Chapter 1, prior studies often normalize the raw signals, so it was believed that finding the non-normalized maximum value per signal segment may be a useful feature, but would require some form of calibration to be useful in an application over a range of different mice.

The fourth and final feature that was calculated from the raw signal was the peakedness of the raw signal. The term "peakedness" is introduced here to describe a simple concept in the signals. In some behaviors, there is produced a high rate of change between signal values. The value itself is calculated as the sum absolute values of the gradient of the signal, or as

$$Peakedness = \sum_{n=0}^{N-1} |x[n] - x[n-1]| \tag{3.7}$$

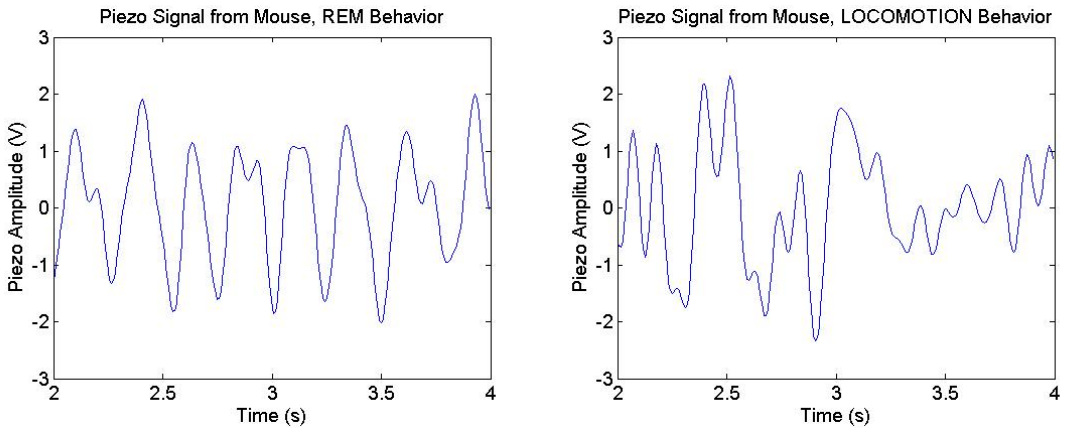
where *N* is equal to the total number of samples in the signal, and $x[n]$ is the normalized signal vector. It will be seen that this feature proved to be one of the more powerful features in the classifiers described below; however, this feature is sensitive to the bandwidth of the signal conditioning filters. For rhythmic behaviors like sleep, the peakedness of the signal is lower than that of the higher activity behaviors. The feature is performed on the normalized signal in an attempt to eliminate the variation in signals due strictly to amplitude, which could be indicative of mice of differing weights, calibration differences, or a combination of the two.

3.6 Power Spectrum Features

Before any power spectrum features were obtained, the frequency values of the power spectrum function p(f) were limited to 0.5 Hz to 11 Hz, since all values above were filtered and suppressed. The values were converted to decibels, according to

$$P(f)(dB) = 20 \log_{10} p(f) \tag{3.8}$$

where P(f) represents the energy values in dB.



(a) (b)

Figure 3.4 Examples of normalized pressure signals. (a) Mouse is in REM sleep. (b) Mouse is moving across cage floor. The peakedness value in the REM segment is 177.97, while the value in LOCOMOTION is 148.56.

The first feature calculated from the power spectrum was the peakedness of the values in the power spectrum. This feature was valuable in discerning behaviors with strong periodic components from those with less periodicity. A higher peakedness here would usually indicate a strong fundamental frequency component, and at times, harmonics.

Here, it is useful to define a feature related to the peakedness, called "deviatedness." Like the peakedness, the deviatedness measures element-to-element variation in a signal. However, it captures a slightly different value by recording the standard deviation of the magnitude of the gradient of a signal, or,

$$Deviatedness = \sqrt{E[P_D^2] - (E[P_D])^2} \tag{3.9}$$

Where $P_D = |P[f] - P[f-1]|$, the gradient of the PSD $P[f]$. The feature's value lies in its ability to detect the variability, or lack thereof, in element-by-element transitions. In

terms of the PSD, the feature detects the variation in amplitude changes from frequency to frequency. If this value is large, it can be assumed that there is a large amount of energy at specific frequencies with respect to the rest of the spectrum, which could indicate breathing activity, or even detect the differences in breathing activities in sleep.

Another feature tested was the kurtosis of the PSD. The kurtosis is defined as the fourth central moment divided by the variance squared minus three, or,

$$Kurtosis = \frac{\mu_4}{\sigma^2} - 3 \tag{3.10}$$

where μ_n is the nth central moment and σ^2 is the variance of the signal. The n^{th} central moment of a signal is defined as the expectation of a signal minus the signal's mean raised to the n^{th} power, or,

$$\mu_n = E[(x(n) - E[x(n)])^2] \tag{3.11}$$

It was proposed that the kurtosis of the PSD could be beneficial in classifying behaviors with strong fundamental frequency peaks from those without, since the kurtosis is lower for signals that appear to have a stronger peak than those without.

The skewness of the PSD is a similar feature to the kurtosis in that it captures characteristics of the waveform in general. The skewness is defined as the third central moment divided by the standard deviation cubed, or,

$$Skewness = \frac{\mu_3}{\sigma^3} \tag{3.12}$$

While the kurtosis of a data segment quantifies the spread of the data from the mean, the skewness measures the asymmetry of data with respect to the mean. It was hypothesized that rhythmic signals indicative of a sleeping or resting mouse would be easily

differentiable from the active mouse, since the spread of energy for the active mouse is much more uniform for the frequencies of interests, leading to a smaller skewness.

The harmonic power ratio (HPR) was calculated as another feature from the power spectrum. As it takes a strong periodic signal to create harmonics, it was believed that the HPR could be yet another useful feature in differentiating the periodic behavior signal segments from the aperiodic behavior signal segments. The HPR calculation follows below as

$$HPR = \frac{\sum_{i=2}^{N} P_i}{P_1} \tag{3.13}$$

where i refers to the ith harmonic frequency and N is the highest harmonic frequency of interest. Since the breathing frequency range is 1.5 to 4.5 Hz, it was sufficient to use N = 4 as a maximum harmonic value to test.

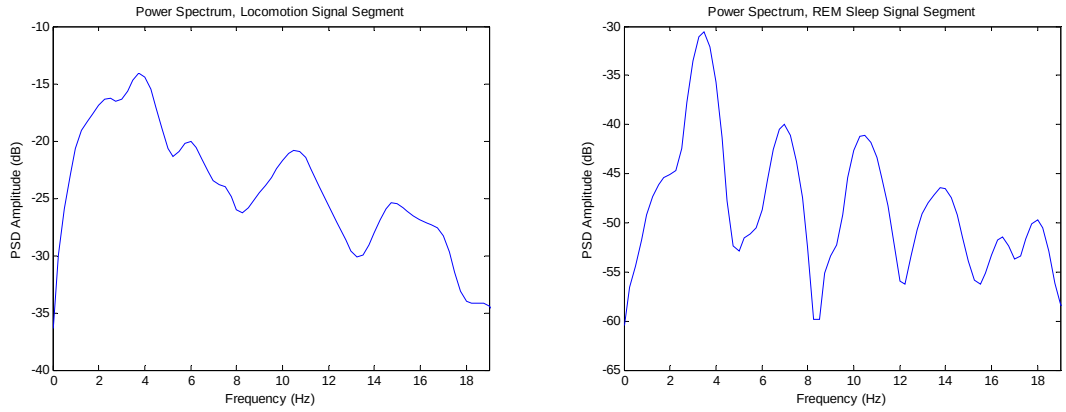
The fourth and fifth features computed in the power spectrum were the height and location of the frequency of the highest energy component. It was believed this feature could be useful in discerning NREM and REM sleep, since the two can have different breathing frequencies. It was also believed that using the max PSD height could be a useful feature in detecting active behaviors with periodic components, such as GROOM, in which the mouse can at times perform a vigorous cleaning motion at a higher frequency that other behaviors do not exhibit.

3.7 Autocorrelation Features

The autocorrelation (AC) was another algorithm that produced valuable feature statistics for the classifier system. The most valuable feature that was found in the AC domain was by taking the standard deviation of the positive lags in the AC as such,

$$\sigma(r) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r(i) - \mu)^2} \tag{3.14}$$

where μ is the mean of the AC function $r(t)$, N is the total number of samples in $r(t)$, and $\sigma(r)$ is the standard deviation of the AC function. This was a feature set that was



(a) (b)

Figure 3.5 Power spectra of (a) LOCOMOTION behavior piezo signal, and (b) REM sleep piezo signal, right. Notice the higher energy at all peaks in the LOCOMOTION signal, but the greater peak definition in the REM sleep signal.

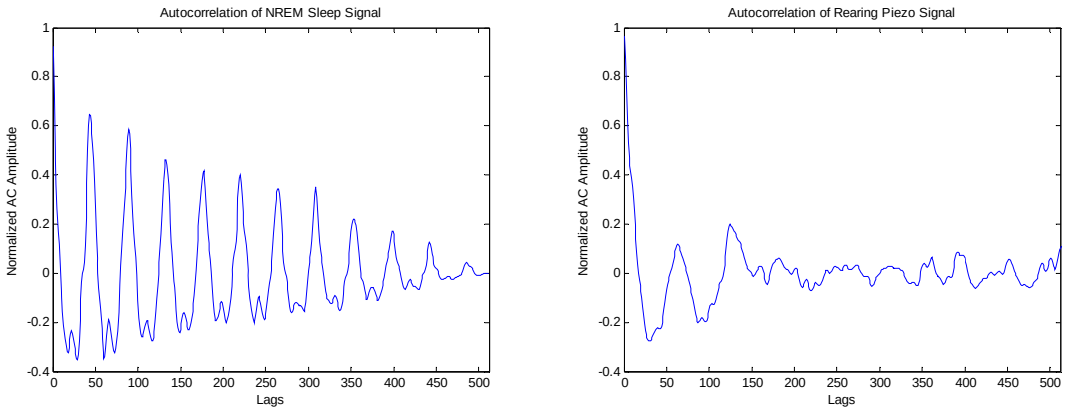
sensitive to periodic signals, and behaviors that produced signals of stronger periodicity were often easily separable using this feature alone. In order to limit the feature to analysis of only the frequencies of interest, only lags corresponding to frequencies between 0.5 Hz and 11 Hz were used in the feature computation.

The Teager energy was computed for the AC as well as the raw signal. It was believed that a signal with less Teager energy in the AC would be more likely to be an active signal, and thus it could be used in the initial sleep and wake classification step, or

in a later step, as in classifying quiet rest from the more active wakeful behaviors. The peakedness feature was also computed for the AC and served a similar purpose.

The peakedness feature was computed on the autocorrelation values using the same function as above. It was hypothesized that the peakedness would detect the peaks in the autocorrelation signal associated with a rhythmic signal. A higher value would indicate the subject was engaging in predictable manner, likely representing sleep or inactivity.

For the final feature that exploited the autocorrelation domain, the value of the standard deviation of the autocorrelation was divided by the deviatedness in the peak heights in the autocorrelation signal. For breathing signals, the autocorrelation signal is repetitive with peaks occurring at regular intervals. The deviatedness of that signal would thus produce a very low value, as compared with the AC from an observed wake signal. In a regular behavior like sleep, therefore, dividing the relatively higher value of the standard deviation of the AC by the low valued deviatedness results in a very high value for the feature, while an irregular behavior records a much lower value, since the numerator value is smaller and the denominator is higher, relative to the regular behavior.



(a) (b)

Figure 3.6 Comparison of autocorrelations of differing behavior signals. (a) NREM Sleep signal. (b) Rearing piezo signal. Notice the much higher standard deviation in the NREM signal, and the greater number of peaks in the REAR signal.

3.8 Signal Envelope Features

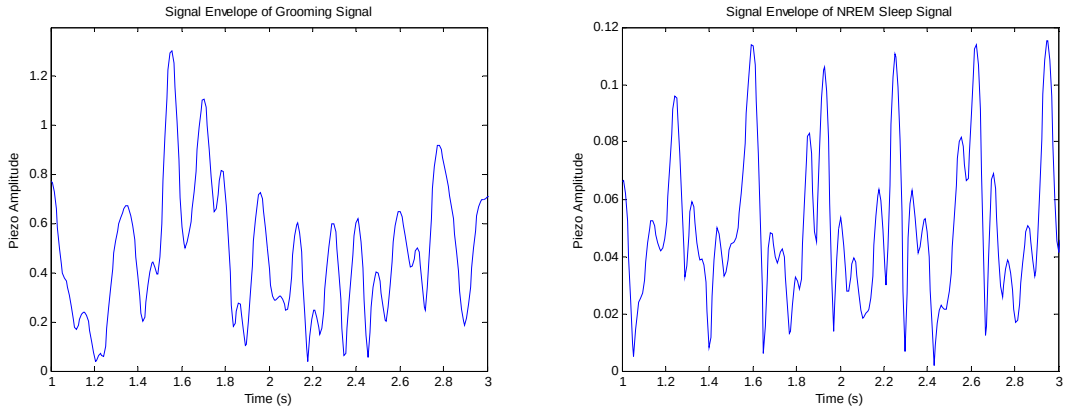
The final domain used to find features was that of the signal envelope. In general, signals with more variable peak heights relate to more active behaviors, thus leading to easier separation of low- and high-activity behaviors. Local maxima were found in the signal envelope, and three features were calculated using these maxima.

The first feature calculated using the maxima in the signal envelope was the standard deviation of the peak heights. Signals with greater variation in peak heights would correspond to active behaviors, and have higher standard deviation values than those of lower activity behaviors.

Similarly, the second feature tracks the range of heights found in the signal envelope. Periodic behaviors should have a very small range in envelope peak heights while aperiodic behaviors like LOCOMOTION tend to have large peak height ranges, especially considering the sometimes strong strikes of the piezo sensor made in this high activity behavior.

Another signal envelope feature tracks the rate of change of the local maxima in the envelope signal. If the differences in values from peak to peak are small, it is reasoned that behaviors have more energy located at one frequency than any other. Thus these signals are more likely to be breathing signals, indicating a state of low activity where the mouse is resting on the cage floor.

The final signal envelope feature is the 20th percentile value for ranked values in the envelope peak heights vector. The motivation for this feature is that, if the values for different behaviors are sorted from least to greatest value, the distributions of values will be characteristic to the individual behaviors, and that a certain percentile value will represent the greatest level of separation in the two classes. The 20th percentile value in particular was chosen for its level of separation in multiple cumulative distribution functions for the behaviors in question.



(a) (b)

Figure 3.7 Signal envelopes. (a) Grooming behavior signal envelope. (b) NREM sleep signal envelope. Notice the greater range of peaks and standard deviation of peaks in the GROOM envelope, left, as opposed to NREM sleep.

Chapter 4 – Classifier Design and Performance

4.1 Introduction

The hypothesis of this work was that using a decision tree based classifier system, previous work could be improved by classifying successfully some the sub-states of sleep, offering a more in depth look into what the piezo pressure sensors generated by the mice subjects can tell us about behavior. This chapter will detail the over-arching view of the decision tree classifier and explain the features used for every stage, and how they affected the system by effectively separating behaviors in the feature space. The chapter is organized in a manner such that each classifier is described in its own subsection, following the system overview section below.

4.2 System Overview

Figure 4.1 below shows the decision tree implementation of the classifier system. Each column shows a new layer of classification and indicates a different classifier has been used to reach that step. While many binary classifiers were designed and tested, it would have been unreasonable to exhaustively test every possible binary system using the 8 states of behavior defined above. Instead, the decision tree follows an intuitive line of thought, and within that line, potential variations were tested to find the optimal systems there within.

The first stage of classification follows previous works and separates sleep and wake behavior. This step as the first step is somewhat obvious due to the radical difference between the two states, where sleep signals are low amplitude and periodic due to the principle breathing signal component, with active behaviors less predictable in amplitude and periodicity. Since this classifier was the most accurate classifier in its own right, of all those tested, it was

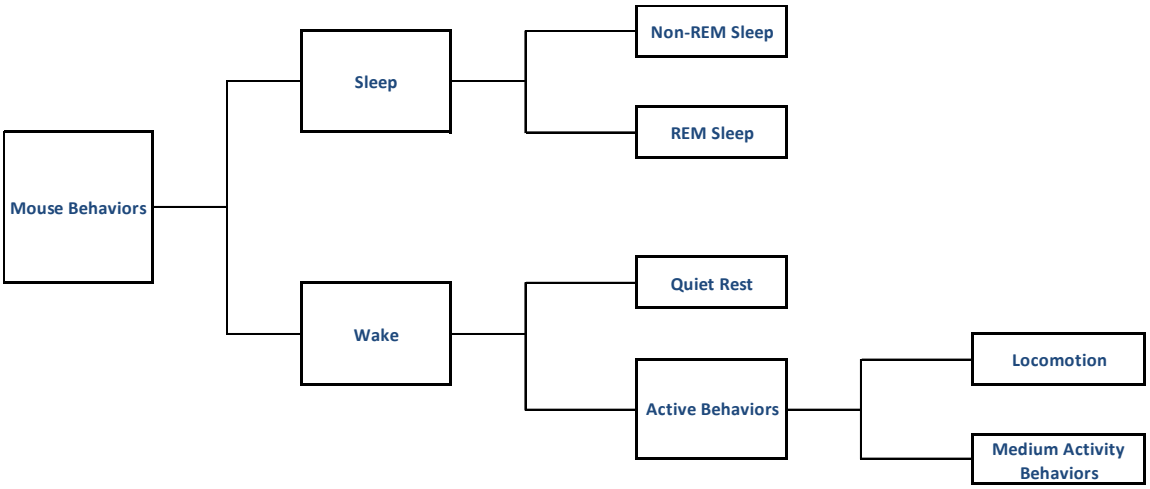


Figure 4.1 Overall decision tree system. Vertical columns represent stages in the classification system.

logical to make it the first step. Errors propagate through the classifier system, so limiting these errors in the early stages of any decision tree system is an important issue.

The next step in classification was to differentiate the substates of sleep and wake. The substates of sleep were chosen to be NREM and REM sleep, while the substates of wake were chose to be "active behaviors" and "quiet rest." The first step wake classification logically classified quiet rest first, as it was the lowest activity wake behavior and was easy to differentiate thanks to its periodic breathing signature. This state was not confused with sleeping behaviors thanks to its occasional spikes which occurred when the mice shifted in some way, and to its slightly differing breathing rate.

At this point, the active behaviors were classified again, this time into the "LOCOMOTION" and "medium activity" substates. Medium activity substates included EAT, DRINK, REAR, and GROOM. Various other classifiers were tested at this step, such as

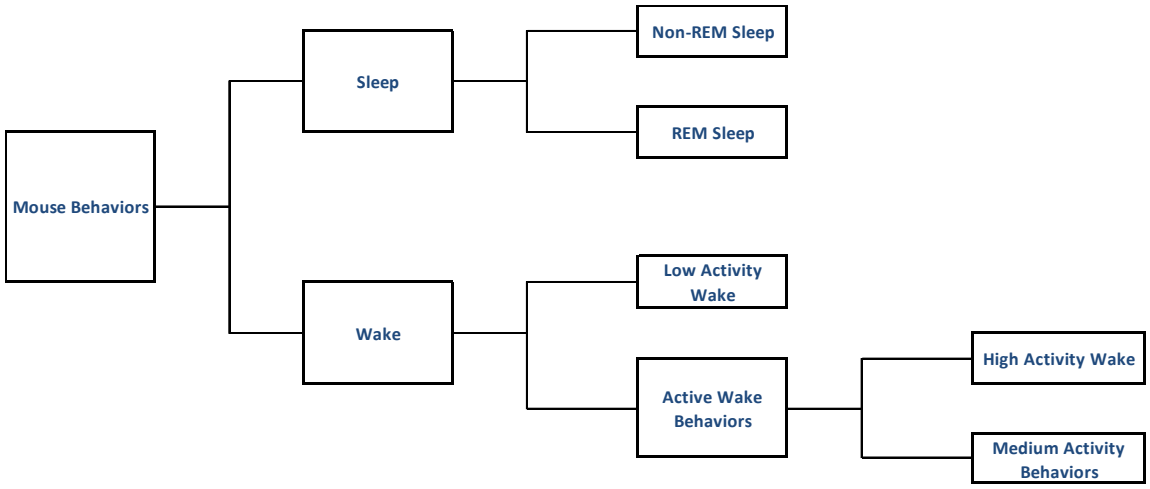


Figure 4.2 Alternative representation of the decision tree classifier system.

classifying feeding activities from non-feeding activities, but none yielded better results than the LOCOMOTION classification. An analogous way to look at the classifier follows below in Figure 4.2.

4.3 Methods and Testing

Some discussion is necessary here to describe the testing method for the classifiers. All classifier training and testing was computed in Matlab. Data were organized into matrices based on the behavior they were observed in. Each behavior matrix had a column dimension size equivalent to the sampling rate used to obtain the data multiplied by the number of seconds that the recording lasted. The row dimension size was equivalent to the number of behavior segments observed; in other words, if the length of the REAR behavior segments were recorded at 8 seconds, and 62 of these behavior segments were observed, the matrix size would be 62 by 1024 samples.

Once a feature statistic was hypothesized to be a valuable classification metric for a set of behaviors, the Fisher's linear discriminant [1] was calculated for that feature to find the level of separation it offered. This calculation was not infallible, but offered a

good idea for how valuable a feature could be. The Fisher's linear discriminant equation follows as

$$\Delta F = \frac{|\mu_1 - \mu_2|^2}{\sigma_1^2 + \sigma_2^2} \tag{4.1}$$

where μ_c is the mean of the feature for class c, σ_c^2 is the variance for class c, and ΔF is the Fisher's linear discriminant value. It was found that values above 0.3 meant a feature could be useful in the classifier, and values of 0.8 or more were almost certainly valuable.

Mice behavior data was available in one of two formats, and unless otherwise noted, all analysis was conducted in Matlab. In the first, the NREM, REM, and WAKE data that was recorded by implanted EEG/EMG electrodes, and scored by trained scorers, was the main source of data for the sleep/wake classification step and the REM/NREM step. The EEG and EMG signals for a given mouse were recorded concurrently with the Piezo pressure sensor readings in Sirenia Sleep. Trained scorers divided the concurrent signals into 4-second epochs and labelled each Piezo segment as either NREM, REM, or WAKE . These scores were provided in a comma separated values (CSV) format for each respective mouse.

A script was written to organize the Piezo data into behavior matrices of 8-second length. The Piezo data was first, however, filtered using a finite impulse response (FIR) bandpass filter (Order = 512, Lower cut-off = 0.5 Hz, Upper cut-off = 11 Hz) to eliminate high frequency noise unrelated to the signals of interest, and to eliminate the DC bias. Each time concurrent behavior segments were discovered in the Piezo data based on the labels, the two concurrent segments were stored as an 8-second segment,

and the next segment was considered. This was repeated for all data for each mouse to store the data.

The second format for the data involved the visual scoring of WAKE behaviors, including GROOM, REAR, EAT, DRINK, LOCOMOTION, and quiet rest. This data was available from visual scorers in Excel (Microsoft, Redmond, WA), and a similar method to that above was used to capture those behaviors that occurred for 8-seconds to store in matrices.

Once the behavior matrices were organized, another script was used to loop through each mouse's matrix and calculate for each segment all the features of interest described above. This data was stored in feature matrices for each mouse.

For the actual classification, a script was written to train and test an LDA classifier using the feature matrices obtained above. The script was structured such that an arbitrary number of Monte Carlo simulations could be tested; generally, 30 simulations were performed for each classifier. The training set was populated by random sampling of the feature matrices for whichever two classes were of interest. Mean feature templates were computed, and a set of linear Fisher weights was obtained. These weights were applied to randomly sampled test feature vectors from each class, chosen to be mutually exclusive from the training set. To find the decision threshold resulting in optimization of classifier accuracy, 200 thresholds were tested for each classifier, and that which yielded the minimum error was chosen as the decision threshold. In general, the calculation for the minimum error assumed equal probability of each class occurring; however, in classifiers where the two class probabilities were unequal, the calculation for the minimum error weighted the more likely class sensitivity more heavily. In other

words, if a class A occurred 85% of the time and a second class, B, occurred the other 15% of the time, the decision threshold was chosen so that the sensitivity of A was higher, since the same percentage of misclassifications of A as B would mean many more raw segments misclassified from A as class B than vice versa. In a case like this, the accuracy of the system would track closely the sensitivity of class A.

The classification function calculated not only the sensitivity rates for each class and the overall system accuracy, but also produced the sets of Piezo segments for each classified class.

Two methods were employed for finding the best feature set for each classifier. The first method was used as a quick method for reducing the feature set for an individual classifier system. It involved testing N-1 permutations of the feature set, and finding the subset of features that produced the lowest error rate in classification. For example, if features 1, 2, and 3 were the original set, then this method would test 3 systems, utilizing features 1 and 2, then 2 and 3, and finally 1 and 3, and would report the lowest error rate generated using those subsets. If one of the subsets had the same or lower rate of incorrect classifications, the process was repeated on that subset. The process would be repeated until the feature set was minimized; i.e., when further reducing the number of features would adversely affect the error rate.

The second method of testing was an exhaustive testing of every possible permutation of feature subsets, and reporting the lowest error rate generated by some set of features. This method has the benefit of yielding the best possible classification rate and finding the features that generate that rate, while the first method uses only marginal improvements and assumes that in every case, the marginally better sets will yield the

best sets overall, which is not necessarily true. The problem with the exhaustive method, however, is the sheer amount of time necessary to run the classifier tests. With 16 features, if every possible value of permutations of every feature length were tested, 65,535 classifiers would need to be trained and tested, and each of those, 200 times, resulting in a total of 13 million systems. The exhaustive process takes hours to test all features for one system, and was not prudent to use in every case.

Usually, during exploratory testing, method 1 would be used. It might be hypothesized that classifying LOCOMOTION and GROOM behaviors from feeding and REAR behaviors might be possible, so method 1 would be used to marginally reduce the feature set. If an overall error rate was found below a certain acceptable value, then the exhaustive method would be employed to further clarify the system values. Generally this step was not necessary, however, as most classifiers tested yielded poor results from method 1.

4.4 SLEEP/WAKE Classifier

The feature set used to classify sleep and wake behaviors follows below in Table 4.1. The sleep state was defined as containing REM and NREM segments, while the WAKE state contained all other substates of behavior. The set below differs from that found by Donohue, et al, [2] offering another approach to Sleep/WAKE classification. It also exemplifies the robustness of LDA for the first classification step, as multiple approaches reach similar success rates. The accuracy of classification for all mice in differentiating the two behaviors was found to be 90%, with sleep sensitivity being 89.3% and WAKE being 91.2% comparable to results reported by Donohue, et al.

Table 4.1 SLEEP/WAKE classifier feature set.

Sleep vs WAKE Classifier Features		Fisher's Linear Discriminant
1	Standard Deviation (Autocorrelation)	1.6539
2	Skewness (PSD)	1.6077
3	Peakedness (Envelope Heights)	1.1591
4	20% Ranked Value (Envelope Heights)	0.2162

The first and most valuable feature in this classification step was found to be the standard deviation of the autocorrelation (AC). As discussed above, the AC was normalized and the standard deviation of the values corresponding to positive lags was computed. Signals that are highly periodic result high periodic self-correlation values, which in turn yield high standard deviation values. The more aperiodic WAKE state signals gave much lower values for this feature, since there was little correlation from time instance to time instance. This feature highly weights periodicities in the classification process, and further helps to separate WAKE and sleep in the feature subspace.

The second feature exploited the difference in skewness in the power spectrum to separate SLEEP and WAKE signals. In the case of mouse behavior analysis via Piezo signal acquisition, if the power spectrum (PSD) is treated as a probability density function, the skewness can be thought of as quantifying how much energy is focused in the frequencies of interest. For a mouse in the SLEEP state, most of the energy in the PSD is centered on the breathing rate, around 3 Hz. This would give the distribution a much longer "tail," thus increasing the skewness value. In the WAKE states, the energy is less focused on any frequency, resulting in a smaller tail and a smaller skewness.

The third feature in the SLEEP/WAKE classifier is the peakedness of the peak heights in the envelope. The peakedness was a valuable feature here due to its ability to

detect a lack of periodicity in the signal. In the aperiodic WAKE signal, the distribution of peak heights in the envelope of the signal was unpredictable and often exhibited great variation from peak to peak. In the periodic SLEEP signals, the peak heights were much more uniform, leading to a lower peakedness value.

The fourth feature was designed to exploit the perceived differing distributions in the signal envelopes. If the peak heights in the signal envelope are sorted from lowest value to highest value, the values resemble a cumulative distribution function (CDF). Since the WAKE signal envelope is much more variable, the resembled CDF should look much different than that of the SLEEP's at various percentiles, thus producing a separable metric.

Table 4.2 shows the accuracy and class sensitivities, with associated 95% confidence intervals, from the SLEEP/WAKE classifier. For each mouse, 200 Monte Carlo simulations were conducted for the classifier using different training and testing sets in each case. From that, a mean accuracy for the system and class sensitivity for each class was obtained, with associated standard deviations. The resulting mean values for each mouse were then used to compute a confidence interval for the mean system accuracy and class sensitivities across all mice. The goal was to provide a general view for how the classifier worked for different mice over a large amount of time. This method was used to calculate the results for the SLEEP/WAKE and NREM/REM classifier systems. Since observation data of active behaviors was available for only one mouse, the confidence interval values reported for the REST/Active WAKE and LOCOMOTION/Medium Active WAKE classifiers following are the result of the 200 Monte Carlo runs for that mouse.

Confidence intervals for the N = 20 mice were computed using the t-statistic, given by

$$CI = \bar{x} \pm t_{crit} \frac{s}{\sqrt{N}} \tag{4.2}$$

Where CI is the 95% confidence interval, $\bar{x}$ is the mean across all mice, t_{crit} is the t-statistic based on N-1 degrees of freedom and $1 - \alpha = 0.975$, and s is the sample standard deviation for the means for all mice.

Table 4.2 SLEEP/WAKE classification results.

SLEEP vs WAKE Classifier Results

	Accuracy	Std Dev	95% Confidence Low	95% Confidence High
System	90.27%	0.55%	90.01%	90.53%
	Sensitivity	**Std Dev**	**95% Confidence Low**	**95% Confidence High**
SLEEP	89.27%	1.16%	88.72%	89.81%
WAKE	91.17%	0.93%	90.73%	91.60%

Table 4.3 presents the system accuracy for each subject based on the respective 200 Monte Carlo simulations, and using the produced standard deviations a 95% confidence interval as described above in the t-statistic computation. For all subjects, the system accuracy was over 86%, and standard deviations from Monte Carlo simulation to simulation were low, above 3% only in one case. This lack of variation in accuracy leads to the conclusion that this classifier is a robust means for discerning SLEEP and WAKE.

Table 4.3 SLEEP/WAKE system accuracy values for each subject, classified by LDA.

	SLEEP/WAKE Accuracy		95% Confidence Intervals	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	91.39%	0.36%	91.34%	91.44%
UK2	90.12%	0.38%	90.06%	90.17%
UK3	92.67%	0.67%	92.58%	92.76%
UK4	90.94%	0.41%	90.89%	91.00%
UK5	90.53%	0.57%	90.45%	90.61%
UK14	89.74%	0.77%	89.64%	89.85%
UK17	93.39%	0.45%	93.32%	93.45%
UK18	86.76%	0.76%	86.66%	86.87%
UK19	86.63%	0.65%	86.54%	86.73%
UK20	88.22%	0.56%	88.14%	88.30%
UK21	89.66%	0.86%	89.54%	89.77%
UK22	91.46%	0.42%	91.40%	91.52%
UK24	90.48%	0.74%	90.38%	90.58%
UK26	89.58%	0.59%	89.50%	89.66%
UK27	90.70%	0.45%	90.64%	90.77%
UK29	86.25%	1.15%	86.09%	86.41%
UK30	91.99%	0.65%	91.89%	92.08%
UK31	89.86%	1.13%	89.71%	90.02%
UK33	88.61%	0.62%	88.52%	88.69%
Tom	92.03%	0.43%	91.97%	92.09%
Totals	90.05%	1.97%	89.13%	90.97%

Table 4.4 SLEEP sensitivity values for each subject, classified by LDA.

	SLEEP Sensitivity		95% Confidence Intervals	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	91.61%	1.10%	91.46%	91.76%
UK2	92.56%	1.74%	92.32%	92.80%
UK3	93.75%	1.22%	93.58%	93.92%
UK4	90.22%	1.00%	90.08%	90.35%
UK5	91.32%	1.11%	91.16%	91.47%
UK14	90.50%	1.42%	90.31%	90.70%
UK17	95.04%	0.55%	94.97%	95.12%
UK18	90.43%	0.70%	90.34%	90.53%
UK19	87.13%	1.42%	86.94%	87.33%
UK20	90.24%	1.44%	90.04%	90.44%
UK21	92.87%	0.91%	92.75%	93.00%
UK22	91.38%	1.04%	91.24%	91.52%
UK24	93.16%	0.82%	93.04%	93.27%
UK26	89.43%	0.75%	89.33%	89.53%
UK27	92.40%	1.00%	92.26%	92.54%
UK29	91.22%	1.48%	91.02%	91.43%
UK30	93.63%	0.98%	93.50%	93.77%
UK31	93.77%	1.35%	93.58%	93.96%
UK33	93.44%	0.87%	93.32%	93.56%
Tom	94.46%	0.82%	94.34%	94.57%
Totals	91.93%	1.94%	91.02%	92.84%

Table 4.4 presents the SLEEP sensitivity means, standard deviations, and 95% confidence intervals for each subject, and Table 4.5 presents the associated values for WAKE sensitivity. In both SLEEP and WAKE, there was little variation from the mean for all subjects leading to the conclusion that the classifier is robust.

Table 4.5 WAKE sensitivity values for each subject, classified by LDA.

	WAKE Sensitivity		95% Confidence Intervals	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	91.18%	0.91%	91.06%	91.31%
UK2	87.31%	1.66%	87.08%	87.54%
UK3	91.52%	1.29%	91.34%	91.69%
UK4	91.63%	0.65%	91.54%	91.72%
UK5	89.40%	1.13%	89.24%	89.56%
UK14	88.86%	1.02%	88.71%	89.00%
UK17	91.11%	1.13%	90.95%	91.26%
UK18	84.35%	1.38%	84.16%	84.55%
UK19	86.11%	1.74%	85.87%	86.35%
UK20	87.17%	1.33%	86.99%	87.36%
UK21	87.79%	1.61%	87.56%	88.01%
UK22	91.54%	1.13%	91.38%	91.70%
UK24	86.82%	1.71%	86.58%	87.06%
UK26	89.71%	1.27%	89.53%	89.88%
UK27	89.26%	1.18%	89.09%	89.42%
UK29	80.85%	1.82%	80.60%	81.10%
UK30	90.03%	1.35%	89.85%	90.22%
UK31	85.13%	1.79%	84.88%	85.38%
UK33	83.45%	1.24%	83.28%	83.62%
Tom	88.56%	0.88%	88.43%	88.68%
Totals	88.09%	2.98%	86.69%	89.48%

4.5 NREM/REM Classifier

The feature set used to classify non-rapid eye movement sleep and rapid eye movement sleep is plotted below in Table 4.6. Unlike in the case of sleep and wake, the differences between REM and NREM, in the Piezo pressure signal, are much subtler, and the Fisher's values in the table illustrate this. No single value is substantially higher than 0.5, and the lowest values are minuscule. However, when this feature set is used, the accuracy in the classification of the substates of NREM and REM sleep on average for all mice is 81%, while NREM sensitivity is 84% and REM sensitivity is 64%. It should be noted that these values correspond to the case where ideal classification of sleep and

wake has occurred; i.e., all NREM and REM sleep segments exist in the set to be classified, and no other behavior segments exist in said set. As always it is important to remember that NREM and REM segments were scored by a trained EEG scorer, so the detections should be thought of as successfully matching what the trained scorer evaluated, and not necessarily the true substate values.

Table 4.6 Feature set used to classifiy REM and NREM substates of sleep, with Fisher's Linear Discriminant values.

NREM vs REM Classifier Features		Fisher's Linear Discriminant
1	Peakedness (Normalized Signal)	0.1025
2	Teager Energy (AC)	0.5036
3	St Dev (AC) / Deviatedness (AC Peak Heights)	0.4647
4	Peakedness (PSD)	0.0112

The Teager energy (TE) of the autocorrelation proved to be the most valuable feature in the NREM/REM classification step. REM sleep is the less periodic SLEEP state and thus has less pronounced peaks in its autocorrelation on average with respect to NREM sleep segments. The TE operator detects the greater amplitudes in the NREM segments and reports a higher energy on average for these segments than REM. However, Teager energy places more weight to energy present at higher frequencies, and in the AC, the REM segments generally exhibit more energy at those frequencies. Thus, the values for TE for REM can be higher as well, leading to a poorer separation than is desired.

The second most valuable feature was originally two separate features, later combined to accentuate the difference between REM and NREM autocorrelations. As has been shown above, taking the standard deviation of the AC results in a higher value for

more periodic signals, and is thus a useful feature in discriminating those from aperiodic waveforms. However, REM sleep is a breathing signal as well, and can have moderately high values for that feature alone. Weighting the standard deviation (AC) feature with the inverse of the deviatedness of the peak heights in the autocorrelation provides resolution between the two classes by further accentuating the less periodic occurrence of peak heights in the AC of REM sleep.

The peakedness in the normalized signal was utilized as a way to further capture the substantial breathing intakes associated with NREM sleep. The feature worked best in cases where the REM breathing was particularly erratic and produced a normalized signal with a lower gradient value.

The fourth and final feature used in this classifier was the peakedness in the power spectrum. It was hypothesized that due to the more variable nature of REM sleep in frequency and amplitude, the peakedness of the power spectrum would result in an effective metric. In fact, as the Fisher's linear discriminant in Table 4.6 shows, this was generally not the case. However some benefit is added by using the feature from the PSD; this set of 4 features consistently produced the best separation of all feature set combinations possible.

As above, 95% confidence intervals were computed for means across all mice of system accuracy and class sensitivities. The results appear in Table 4.7.

Table 4.7 NREM/REM classification results.

REM vs NREM Classifier Results

	Accuracy	Std Dev	95% Confidence Interval	
			Low	High
System	80.86%	3.03%	79.44%	82.28%
	Sensitivity	**Std Dev**	**95% Confidence Interval**	
			Low	**High**
NREM	83.64%	3.31%	82.09%	85.19%
REM	64.21%	10.98%	59.07%	69.35%

Table 4.8 presents the system accuracy for each subject based on the respective 200 Monte Carlo simulations, and using the produced standard deviations a 95% confidence interval as described above in the t-statistic computation. For all but one subject, the system accuracy was over 75%, and standard deviations from Monte Carlo simulation to simulation were low, above 3% only in one case. This lack of variation in accuracy leads to the conclusion that this classifier is a robust means for discerning REM and NREM.

Table 4.8 Mean NREM/REM classification accuracy for each subject.

	NREM/REM System Accuracy		95% Confidence Interval - Mean Accuracy	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	83.71%	1.42%	83.51%	83.91%
UK2	82.96%	1.13%	82.80%	83.12%
UK3	81.83%	1.80%	81.59%	82.08%
UK4	80.57%	1.43%	80.37%	80.77%
UK5	79.72%	1.89%	79.46%	79.98%
UK14	83.95%	2.21%	83.65%	84.26%
UK17	83.82%	1.19%	83.66%	83.99%
UK18	72.83%	1.06%	72.68%	72.98%
UK19	77.75%	1.25%	77.58%	77.92%
UK20	82.88%	2.32%	82.56%	83.21%
UK21	75.18%	3.13%	74.74%	75.61%
UK22	80.65%	1.63%	80.42%	80.88%
UK24	82.98%	2.28%	82.66%	83.29%
UK26	79.77%	2.12%	79.47%	80.06%
UK27	77.58%	1.85%	77.32%	77.83%
UK29	81.58%	1.65%	81.35%	81.81%
UK30	83.57%	2.01%	83.30%	83.85%
UK31	80.71%	1.47%	80.51%	80.92%
UK33	82.07%	1.65%	81.84%	82.29%
Tom	83.11%	1.24%	82.94%	83.29%
Totals	80.86%	3.03%	79.44%	82.28%

Table 4.9 presents the NREM sensitivity means, standard deviations, and 95% confidence intervals for each subject, and Table 4.10 presents the associated values for REM sensitivity. While the NREM sensitivity values were generally classified well and had a low level of variation from the mean, REM sensitivity values varied greatly. Mean-to-mean standard deviation was over 10%, and the 95% confidence interval is much wider. There are a number of factors that could lead to this increased level of disparity. REM is simply a difficult behavior to detect due to its lower level of prevalence and the subtle markers that indicate its distinct breathing signature. It is also simply a more difficult state for human observers to successfully score, and if signal quality in EEG or

EMG signals is poor, the human scoring accuracy will be decreased. It is worth pointing out again here that this classifier system attempts to find agreement with human scorers, which does not necessarily indicate agreement with the actual states of sleep.

Table 4.9 NREM sensitivity values for each subject.

	NREM Sensitivity		95% Confidence Interval - Mean NREM Sensitivity	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	86.65%	2.20%	86.35%	86.96%
UK2	89.74%	1.94%	89.47%	90.01%
UK3	85.45%	2.67%	85.08%	85.82%
UK4	81.91%	2.17%	81.61%	82.21%
UK5	82.34%	2.81%	81.95%	82.73%
UK14	85.89%	2.95%	85.48%	86.30%
UK17	85.50%	1.47%	85.30%	85.71%
UK18	85.61%	2.53%	85.26%	85.96%
UK19	80.85%	2.50%	80.51%	81.20%
UK20	82.89%	2.89%	82.49%	83.29%
UK21	76.39%	4.43%	75.77%	77.00%
UK22	81.42%	2.13%	81.12%	81.71%
UK24	86.14%	2.92%	85.74%	86.55%
UK26	80.32%	2.76%	79.93%	80.70%
UK27	78.00%	2.40%	77.67%	78.33%
UK29	84.57%	2.25%	84.26%	84.88%
UK30	84.85%	2.59%	84.50%	85.21%
UK31	81.70%	1.82%	81.45%	81.95%
UK33	84.57%	2.16%	84.27%	84.87%
Tom	88.07%	1.79%	87.82%	88.32%
Totals	83.64%	3.31%	82.09%	85.19%

Table 4.10 REM sensitivity values for each subject.

	REM Sensitivity		95% Confidence Interval - Mean REM Sensitivity	
Mouse Name	Mean	Standard Deviation	Low	High
UK1	66.36%	3.26%	65.91%	66.81%
UK2	49.57%	3.02%	49.15%	49.99%
UK3	59.45%	3.72%	58.93%	59.96%
UK4	72.05%	3.35%	71.58%	72.51%
UK5	62.59%	4.18%	62.01%	63.17%
UK14	68.18%	3.87%	67.64%	68.71%
UK17	63.89%	2.20%	63.58%	64.19%
UK18	35.93%	3.34%	35.47%	36.39%
UK19	67.76%	2.97%	67.35%	68.17%
UK20	82.81%	4.12%	82.24%	83.38%
UK21	66.63%	6.24%	65.76%	67.50%
UK22	73.10%	3.39%	72.63%	73.57%
UK24	58.60%	3.00%	58.19%	59.02%
UK26	75.55%	3.07%	75.12%	75.98%
UK27	74.40%	2.99%	73.99%	74.82%
UK29	58.06%	3.54%	57.57%	58.55%
UK30	71.64%	4.03%	71.08%	72.20%
UK31	70.10%	3.17%	69.66%	70.53%
UK33	61.40%	3.30%	60.94%	61.86%
Tom	46.10%	3.12%	45.67%	46.53%
Totals	64.21%	10.98%	59.07%	69.35%

4.6 REST/Active WAKE Classifier

This classifier, also described as a low activity WAKE/active WAKE classifier, would be used to classify the WAKE data segments as classified by the SLEEP/WAKE system. The logical grouping for this step was to differentiate the segments where the mouse was quietly resting but awake, since these offer periodic breathing signals that are easily separable from the less periodic active WAKE behaviors. In fact, the feature set relies on features that exploit this very fact. The low active wake subset refers to the behavior substate also known as quiet rest,

Table 4.11 Low activity WAKE versus active WAKE classifier features.

REST vs Active WAKE Classifier Features		Fisher's Linear Discriminant
1	Teager Energy (AC)	0.1821
2	Kurtosis (PSD)	0.1245
3	Peakedness (Envelope Heights)	0.7172

while the active wake subset of behaviors refers to the GROOM, LOCOMOTION, EAT, DRINK, and REAR substates.

The most valuable feature in the set proved to be the peakedness in the Envelope Heights. The characteristic breathing signature produces a low value for the peakedness in the REST substate, in contrast to the erratic active WAKE signals and their greater variation in envelope peak heights.

The Teager energy of the AC offered some separation in the two classes by again capitalizing on the higher energy level in the AC of the less variable breathing rates. Due to shifts in posture as the mouse becomes more comfortable or sniffs the air in the cage, the breathing signal is corrupted by a noise-like component, which reduces the self-similarity of the signal at increasing lags. However, the AC still detects the periodicity belying these motions, and thus will produce a higher value in Teager energy than will the active WAKE behaviors, which have comparatively little self-similarity.

In the domain of the power spectrum, the less periodic active WAKE signals contain less centralized energy in the region of interest, and thus produce a lower value of kurtosis than the REST signals. The feature offered some separation, but was slightly less valuable since in REST there is more energy outside the breathing frequency region, resulting from perturbations of a shifting mouse.

Given the small feature set and the low values for linear discriminants for the respective features, the active WAKE behaviors were classified successfully from low active, REST behaviors. The values presented in Table 4.12 show the classification system accuracy and class sensitivity means assuming the sleep/wake classification step was perfect. Since observation data of the active behaviors was available for only one mouse, the values reported show the average rates for 200 classifiers designed using varying subsets of behavior segments for the training and testing sets. 95% confidence intervals for the N = 200 Monte Carlo runs were computed via use of the z-statistic, as

$$CI = \bar{x} \pm z_{score} \frac{s}{\sqrt{N}} \tag{5.2}$$

Where CI is the 95% confidence interval, $\bar{x}$ is the mean across all mice, z_{score} is the z-statistic based on N = 200 Monte Carlo simulations and $1 - \alpha = 0.95$, and s is the sample standard deviation for the means for all mice.

Table 4.12 REST/Active WAKE classification results.

REST vs Active WAKE Classifier Results

	Accuracy	**Std Dev**	**95% Confidence Interval**	
			Low	**High**
System	91.75%	0.34%	91.70%	91.80%
	Sensitivity	**Std Dev**	**95% Confidence Interval**	
			Low	**High**
REST	93.66%	0.63%	93.57%	93.75%
Active WAKE	90.24%	0.93%	90.11%	90.37%

4.7 LOCOMOTION/Medium Active WAKE Classifier

At this level in the decision tree, it was much less obvious which behavior substate, or set of substates, were most efficiently classified using the information presented. However, since the amount of behavior substates at this point was small enough, exhaustive testing was possible. Every reasonable combination of the five remaining substates – GROOM, LOCOMOTION, REAR, EAT, and DRINK – were tested in a binary classification scheme. To test this, for example, GROOM and LOCOMOTION and REAR were grouped together and classified against the feeding behaviors, to test one possible configuration. Using the feature superset of N total features, all combinations of features were tested exhaustively for each number of features.

The most effective classifier system proved to be a binary classification between high activity, or LOCOMOTION, and medium activity, or REAR, GROOM, EAT, and DRINK, substates. This is likely due to the fact that while in the LOCOMOTION substate, all four of the mouse's feet are on the pressure sensor and the mouse is geospatially moving about the cage. In the medium activity substates, the mice are in the same location in the cage and are on their hind legs only for the majority of the behaviors.

Table 4.13 Linear discriminants for the features in the high active WAKE and medium active WAKE classifiers.

LOCOMOTION vs Medium Active WAKE Classifier Features		Fisher's Linear Discriminant
1	Peakedness (Normalized Signal)	0.7119
2	Skewness (PSD)	0.1852
3	Peakedness (Envelope Heights)	0.7172
4	20% Ranked Value (Envelope Heights)	0.0714

Two features were exceptionally useful in this classifier and revolved around the use of the peakedness measure of differing domains. The peakedness measures of the normalized signal and of the envelope heights were found to be effective for the same reason. The LOCOMOTION substate was by far the most variable state in terms of its pressure signal produced. While in the medium active substates, the mouse is performing a somewhat repetitive motion. For example, while eating the mouse is on its hind legs in one position on the sensor, and either moving towards the food dispenser to get food, or sitting on its haunches eating. DRINK saw the same repetitive motion, as the mice discernibly perturbed the sensor at a higher frequency due to the reaching and licking motions. Grooming, too, was more periodic than the locomotion signal, as the mouse at times vigorously shook while using its paws to clean its fur or tail.

In the normalized signal, the LOCOMOTION signals had a much higher variation and produced a higher peakedness. Likewise, the LOCOMOTION envelope heights were more variable than the medium active WAKE substates and the value for peakedness was again higher.

Less valuable but still useful was the skewness in the PS. The skewness for the LOCOMOTION signals were generally slightly higher since more energy in these signals was concentrated outside the frequencies of highest energy.

Finally, the 20% ranked value in the envelope heights feature added separation in the two states by exploiting the difference in the cumulative distribution functions (CDF) for the two classes. It was assumed that the LOCOMOTION CDF was Gaussian, whereas the medium active WAKE CDF was less normal. Different percentiles were tested but at

20% the difference in the CDFs were greatest, and thus the feature provided useful separation.

Table 4.14 shows the means and confidence intervals for system accuracy and class sensitivities of the high active/medium active behavior classifier branch in the decision tree. The values in the table represent perfect classification in the preceding steps, and are meant to show the value of the features and their ability to classify these behaviors without the errors of the previous stages propagated through. As in the case of the REST/Active WAKE classifier, the z-statistic from Equation 5.2 was used to compute the 95% confidence intervals.

Table 4.14 Classifier results for the high active (LOCOMOTION) and medium active (GROOM, EAT, DRINK, and GROOM) substates.

LOCOMOTION vs Medium Active WAKE Classifier Results

	Accuracy	**Std Dev**	**95% Confidence Interval**	
			Low	**High**
System	78.26%	1.04%	78.12%	78.40%
	Sensitivity	**Std Dev**	**95% Confidence Interval**	
			Low	**High**
LOCOMOTION	80.02%	1.66%	79.79%	80.25%
Medium Active WAKE	77.15%	2.70%	76.78%	77.52%

4.8 Conclusion

The overall decision tree classification scheme is shown below in Figure 4.1. The percentages next to each behavior group show the conditional probability of successful classification. In other words, given the preceding step classified its average rate of

successful classifications, the current step behavior classifies the segments presented and errors propagate through the system. So, in the case of NREM sleep, given all behaviors, 74.79% are classified appropriately as NREM, and 25.21% of those segments classified as NREM are incorrect. Since the NREM step occurs after the SLEEP/WAKE step, some REM segments and some WAKE segments will be classified as NREM sleep, accounting for the 25.21% of the behavior.

Further refinement was attempted but given the feature super set used for the preceding steps, no set was found that could further classify the behavior substates. Groupings were exhaustively tried, and new features were tried as well, but no satisfactory classifiers were developed at this point.

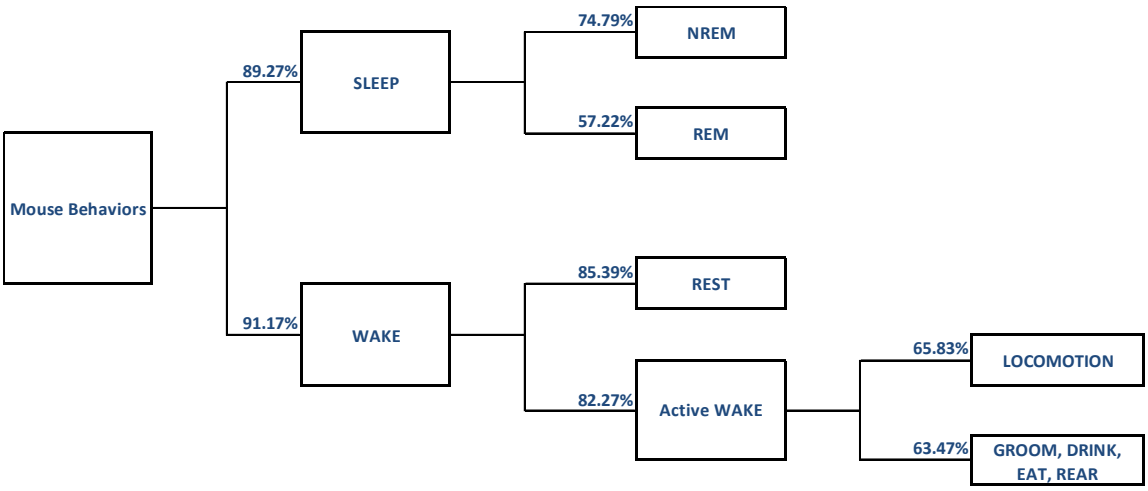


Figure 4.3 Overall decision tree classifier system with percentages of classification successes, including propagated errors from previous steps.

It is worth noting that the SLEEP/WAKE system had a comparable system accuracy to the Piezo-based acquisition and classification system presented by Donohue, et al [2] using a different feature set, lending further validation to the use of LDA and PVDF sensors to detect breathing rates and automatically classify sleep from wake. The

NREM/REM system offers further promise in the automated classification of the substates of sleep based on breathing patterns, with its observed 81% system accuracy, and high level of sensitivities in as a cascaded classifier as in Figure 4.1.

Chapter 5 - Classifier Performance in the Presence of Noise, and Compared to Other Classifiers

5.1 Introduction

As mentioned before, the decision tree classifier system presented was neither the first system tested nor the only configuration tested. In addition to the binary classifiers in the decision tree, 3-state, 4-state, and 8-state classifiers were designed and implemented as well and are presented below to give a basis for comparison to the success rates in the binary system. It was also desired to test the robustness of the decision tree binary system in the presence of differing types of noise and at differing levels of noise. The results of those tests are presented below as well.

There is a low-level baseline noise spectrum associated with the piezo sensor system which at the frequencies of interest is assumed to have a negligible effect on signals. This assumption is generally not a bad assumption, as usually even a low energy signal produced by a sleeping mouse will produce a spectrum on average 15 dB above the noise envelope. There is a peak around 8 Hz in the baseline noise spectrum, to be discussed at greater length below.

Signal-to-Noise Ratios (SNR) are used as the independent variable in the noise tests, and for this application are calculated by

$$SNR = \left(\frac{\sigma_{s,rms}}{\sigma_{n,rms}}\right)^2 \tag{5.1}$$

where $\sigma_{s,rms}$ is the mean root-mean-square value of the behavior signal segments, and $\sigma_{n,rms}$ is the standard deviation of the additive noise signal. Prior to this calculation and the inclusion of the additive noise, the piezo pressure signals were normalized by their behavior's respective mean standard deviation. This formula was ideal as a comparison tool since the noise applied was known and controlled explicitly.

5.2 Noise Testing

5.2.1 Additive Gaussian White Noise

The NREM/REM system was tested for the effects of noise on the success due to the nature of the two signals and their inherent similarities. Both signals feature a low amplitude signal dominated by breathing oscillations, lending to the periodic nature of the signals. The differences in the two states have been described at length above, but it is suspected that the presence of noise will disrupt the classifier's ability to discern the two signals, since the classifier mainly relies on the assumption that REM breathing signals are slightly more variable than NREM breathing patterns. In the presence of white noise, NREM breathing patterns will show greater variability, leading to the disruption of the subtle differences that characterize these two states.

To test the system, a Gaussian white noise signal of length ten times that of the piezo segments was generated using the pseudorandom number generator function randn.m (Matlab). The noise signal standard deviation was normalized to mean standard deviation of the REM and NREM signals. A portion of this noise signal, equivalent to the length of the piezo segments, was removed. This data sequence was used as the simulated noise component and was added element-wise to the NREM and REM segments prior to feature computation. The NREM and REM segments were normalized by the mean

standard deviation of the sleep behavior signals prior to the addition of the noise. The simulated noise component's standard deviation was varied in order to test differing noise levels. Following the feature computation, the classifier was tested as above.

Table 5.1 below shows the success rates for each substate, NREM and REM, in the presence of noise. The first entry showing "inf" SNR indicates the ideal case, where it is assumed the baseline noise is well below the signal energy. Representative figures also follow showing the noise corruption for a REM signal and a NREM signal at various amplitudes, as well as PSD plots for the noise and signal segments.

Table 5.1 Classifier success rates in the presence of white noise, as categorized by signal to noise ratio (SNR).

SNR (dB)	System Accuracy	NREM Sensitivity	REM Sensitivity
No Noise Added	80.95%	83.78%	64.10%
6.02	79.08%	82.57%	61.37%
0	80.51%	84.47%	60.29%
-13.98	76.94%	84.96%	36.39%
-20	79.48%	90.99%	21.21%

As seen, the classifier remains effective up to about an SNR of around 0 dB, and then begins to suffer the effects of the white noise. The NREM sensitivity remains the same and then begins to grow, but the REM segments begin to deteriorate quickly with decreasing SNR. Since NREM is so much more prevalent in sleep than REM, the accuracy remains constant as NREM increases slightly and REM sensitivity decreases drastically. Table 5.2 below illustrates the cause of this deterioration in quality. The Fisher's Criterion was calculated for the five features used in the classifier in two cases; first, assuming no noise is present; and second, with the additive white noise such that the SNR was -3.72 dB. All features suffered drastically in the presence of the noise, with

greatly reduced discriminant values. As higher levels of white noise are added to the signal, the low level variants that characterize and discern REM from NREM are washed out, and the classifier is unable to determine which segments are REM.

Table 5.2 Fisher's Linear Discriminant values in the presence of Gaussian white noise.

Feature	No Noise	Noise
Teager Energy of Raw Signal	0.0568	0.0565
Peakedness of Raw Signal	0.1086	0.0701
Range, Envelope Peak Heights	0.0161	0.0071
Peakedness, Envelope Peak Heights	0.0175	0.0465
Teager Energy Autocorrelation	0.1722	0.0125

5.2.2 Noise at Specified Frequency

In some cases the mouse cage data acquisition system recorded a specific noise peak at the 8 Hz segment. In a few of those cases, this noise peak was strong enough to corrupt the data enough to render the classifier ineffective. Using a slightly different signal acquisition circuit at the DAQ interface, much of this noise was blocked. However, the signals that characterize breathing in mice are so low in amplitude that even small amounts of systematic noise could affect the classification greatly. The robustness of the classifier was tested in the face of a specific noise component to find the maximum tolerable signal to noise ratio for the system. Table 5.3 illustrates the system's response to increasing levels of noise relative to input signal.

The 8 Hz noise component was generated by adding a cosine with frequency $2*\pi*8$ and constant amplitude element-wise to the NREM and REM segments and classifying these new segments. This constant amplitude would be altered from test to test in order to change the signal-to-noise ratios. As above, the SNR was calculated by dividing the signal's standard deviation by that of the noise, and then converting to decibels.

Table 5.3 Classifier success rates in the presence of an 8 Hz noise signal at varying SNR values.

SNR (dB)	System Accuracy	NREM Sensitivity	REM Sensitivity
No Noise Added	80.95%	83.78%	64.10%
6.02	81.53%	85.38%	62.08%
0	79.62%	82.97%	62.68%
-13.98	80.12%	93.46%	12.87%
-20	78.31%	89.23%	23.03%

Referring to Table 5.4, the linear discriminant values for the features again are reduced drastically with a high level of additive noise. However, the Teager Energy (AC) feature is not as adversely affected in this case. As stated above, the Teager Energy metric more heavily weights the energies at higher frequencies. In the case of white noise, there is uniform energy at every frequency, which would cause the TE to be higher for both NREM and REM. With an additive component of only 8 Hz, however, TE detects no noise at higher frequencies and thus is reduced less in a comparable SNR level.

When it is considered that the subtle differences in REM and NREM states are detectable and only slightly reduced in the face of either type of noise at an SNR equal to zero, it must be concluded that the feature set for the classifier is robust. Generally, noise levels recorded are not so high that their affects cannot be filtered easily, or that certain components completely conceal the low-amplitude breathing signal, and in most cases it appears that the NREM/REM classifier is adequate. Since the other classifiers presented in the decision tree are either less subtle in their class differences or have higher amplitude behaviors, it is reasoned that they will be even less susceptible to the effects of noise than the NREM/REM subsystem.

Table 5.4 Fisher's Linear Discriminant values in the presence of 8 Hz white noise, with SNR of -3.92.

Feature	No Noise	Noise
Teager Energy of Raw Signal	0.0568	0.0567
Peakedness of Raw Signal	0.1086	0.0952
Range, Envelope Peak Heights	0.0161	0.0161
Peakedness, Envelope Peak Heights	0.0175	0.0427
Teager Energy Autocorrelation	0.1722	0.0717

5.3 Alternative Classifier Configurations

The following classifier configuration was designed to differentiate more than two classes at once. As such, linear discriminant analysis could not be used. Instead, the minimum distance Mahalanobis distance classifier was employed. The following system is included here to compare the accuracy and sensitivity of a three class minimum distance classifier system to differentiate NREM, REM, and WAKE to that of the LDA binary classification decision tree system presented above.

5.3.1 Three Class System

The first classifier system focuses on the two sleep substates and the general wake state, and could have been used to subvert the need for two steps to classify REM and NREM behavior states. This system could be useful in cases where only NREM, REM, and wake states are desired, as in the case of typical EEG analysis. Separating the states in one step could open the door to real-time classification as well, since the system would require less time than a two-step system.

The program flow is much the same as the binary classifiers presented above. Features were computed for the three classes and stored for use in the classification

algorithm. In the classification algorithm, three feature matrices were generated from random sampling the feature vectors of the respective classes, resulting in three training and three testing matrices. Three training covariance matrices were computed for the Mahalanobis classification step. The first major difference came in the calculation of which class a particular segment belonged to. Each behavior test segment was tested in the template designed for its class, and when the distance calculated was shorter to its particular class than to either of the other two classes, it was stored in a particular matrix. Incorrect classifications were stored in such a way that indicated which class a behavior was incorrectly classified into.

As above, 200 Monte Carlo runs were computed with differing testing and training feature vectors, and sensitivity rates were recorded.

The features that produced the most effective classifiers are reported in the table below. The features were determined by the iterative method used elsewhere of finding the best results from each feature set as determined by subtracting one feature from the superset. The feature set was similar to the SLEEP/WAKE and NREM/REM classifiers described above, in that the most useful features from the two binary classifiers were useful in the three-class classifier, as well.

Table 5.5 Feature set used in the NREM-REM-WAKE classifier.

Feature Set	
1	Standard Deviation (AC)
2	Peakedness (Normalized Signal)
3	Teager Energy (AC)
4	Skewness (PS)
5	Kurtosis (PS)
6	Peakedness (Envelope Heights)
7	80% Ranked Value (Envelope Heights

The number of features was greater than in any of the decision tree classifiers, as expected for the case of splitting three classes. The results are reported in Table 5.2.

Table 5.6 Confusion matrix demonstrating sensitivity rates for the three behaviors.

		Predicted		
		NREM	REM	WAKE
Actual	NREM	92.19%	4.90%	2.91%
	REM	51.45%	23.58%	24.97%
	WAKE	6.32%	2.97%	90.71%

Clearly, the one step classifier is inferior to the cascaded binary classifiers presented in this work. REM sensitivity suffers, as half of the REM sleep state segments were lumped into the NREM class. The overall accuracy of the classifier was fairly high, however, at 91%. The vast majority of segments are NREM and WAKE, so accuracy tends to track closely with the sensitivity of those behaviors. Since REM detection is of

greatest interest, it would be preferable to isolate all SLEEP segments from WAKE before attempting to discern REM and NREM sleep.

Chapter 6 – Conclusions and Future Work

6.1 Conclusions and Future Work

This work presented a decision tree, binary step classifier system for successfully differentiating mouse behavior and substates of those behaviors, using features obtained solely from the pressure signals generated by a mouse as it exhibits these behaviors atop a piezo pressure signal. The system presented was shown to adequately differentiate between sleep and wake, and also illustrated high rates of success in classifying sub behaviors such as non-rapid eye movement sleep and rapid eye movement sleep, quiet rest in the wakeful state, LOCOMOTION, and the remaining medium activity behaviors.

Due to the scope of the work, which involves the design of multiple classifier systems and associated selection of feature sets, a feature superset of 16 statistical metrics was used. These metrics were chosen based on both their intuitive theoretical application to the classification problem as well as their success in differentiating a sub behaviors. These features were chosen from four domains of signal representation in an attempt to add both breadth and depth to the training and testing of the signals. These 16 features were believed to be very good for the current application, but there is no way to say for sure if this was the best set that could have been used. Future studies could further inspect the feature superset here and perhaps add valuable metrics to those presented here. One area of statistics that these may come from are metrics that attempt to quantify the shape of distributions of values, such as skewness, interquartile ranges, etc. These features were tested in the preliminary phase, but that testing was abandoned as there were no clear metrics that were found that worked well enough to be included. This testing was not and could not be exhausted, so there is room to improve there.

Another area of future research could be in the organization of the decision tree. It was shown that an alternative single step classifier could be designed without exhaustive feature testing that classified certain wake behaviors as well as or better than the decision tree. While this approach is difficult to implement since it is hard to find a small, robust feature set that sufficiently and explicitly differentiates each behavior from every other, it could be useful to design a decision tree based system that is not binary at every step, or, to design a classifier that assigns a confidence value to each branch, and selects a terminal node for a segment with the highest sum of the confidences for that trajectory.

The wake behaviors could be more closely scrutinized in future work. It is not clear how certain behaviors generate certain piezo signals, so future work to examine these behaviors could prove to be valuable. If it is known that GROOM, for example, produces a signal of certain power at certain frequencies due to the motions involved in that behavior, then features could be appropriately extracted to isolate the behavior. Some trends were observed in each of the active behaviors, but it would be outside the scope of this work to investigate the physical motion of the mouse and its corresponding pressure signal on a segment by segment basis.

It has been shown that differing strains of mice exhibit differing breathing rates in the phases of sleep, so some work could go to testing this system on many strains of mice, to see if there is any variation in the classification rates. Two strains were used to generate the piezo and EEG signals for this study, and these data were taken together and classified without respect to mouse strain. Since breathing rates differ, REM and NREM sleep classification, which relies on features that extract information about periodicities in

the pressure signals based on breathing rates, could improve or worsen from mouse strain to strain.

Overall, however, this system proved successful in its original goal, and differentiated behaviors based on a decision tree system quite well, especially with respect to determining the phases of sleep appropriately around 80% of the time on average.

References

[1] Mural, R. J. (2002). A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science*, *296*(5573), 1661-1671.

[2] Chung-Chuan, L. (2004). Common scale-invariant patterns of sleep–wake transitions across mammalian species. *PNAS*, *101*(50), 17545-17548 .

[3] Donohue, K. D., Medonza, D. C., Crane, E. R., & O'Hara, B. F. (2008). Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice. *Biomedical engineering online*, *7*(1), 14.

[4] Medonza, D.C. (2006). *Automatic detection of sleep and wake states in mice using Piezoelectric sesnors (Master's thesis).* University of Kentucky, Lexington, KY.

[5] Proakis, J., & Manolakis, D. (2007). *Digital Signal Processing: Principles, Algorithms, and Applications* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

[6] Megens, A. A., Voeten, J., Rombouts, J., Meert, T. F., & Niemegeers, C. J. (1987). Behavioural activity of rats measured by a new method based on the piezo-electric principle. *Psychopharmacology*, *93*(3), 382-388.

[7] Flores, A. E., Flores, J. E., Deshpande, H., Picazo, J. A., Xie, X., Franken, P., ... & O'Hara, B. F. (2007). Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements. *Biomedical Engineering, IEEE Transactions on*, *54*(2), 225-233.

[8] Donohue, K. D., & O'Hara, B. F. (2008, April). SVM for automatic rodent sleep-wake classification. In *Southeastcon, 2008. IEEE* (pp. 581-586). IEEE.

[9] Kjellstrand, P., Holmquist, B., Jonsson, I., Romare, S., & Månsson, L. (1985). Effects of organic solvents on motor activity in mice. *Toxicology*, *35*(1), 35-46.

[10] Marsden, C. A., & King, B. (1979). The use of Doppler shift radar to monitor physiological and drug induced activity patterns in the rat. *Pharmacology Biochemistry and Behavior*, *10*(5), 631-635.

[11] Zeng, T., Mott, C., Mollicone, D., & Sanford, L. D. (2012). Automated determination of wakefulness and sleep in rats based on non-invasively acquired

measures of movement and respiratory activity. *Journal of Neuroscience Methods*, *204*(2), 276-287.

[12] Clarke, R. L., Smith, R. F., & Justesen, D. R. (1985). An infrared device for detecting locomotor activity. *Behavior Research Methods, Instruments, & Computers*, *17*(5), 519-525.

[13] Hilakivi, L. A., & Hilakivi, I. T. (1986). Sleep-wake behavior of newborn rats recorded with movement sensitive method. *Behavioural brain research*, *19*(3), 241-248.

[14] Päivärinta, P., & Korpi, E. R. (1989). Automated method for measuring fighting behavior and locomotor activity of mice. *Physiology & behavior*, *45*(4), 857-860.

[15] Gegout-Pottie, P., Philippe, L., Simonin, M. A., Guingamp, C., Gillet, P., Netter, P., & Terlain, B. (1999). Biotelemetry: an original approach to experimental models of inflammation. *Inflammation Research*, *48*(8), 417-424.

[16] Tang, X., & Sanford, L. D. (2002). Telemetric recording of sleep and home cage activity in mice. *Sleep*, *25*(6), 691-9.

[17] Storch, C., Höhne, A., Holsboer, F., & Ohl, F. (2004). Activity patterns as a correlate for sleep–wake behaviour in mice. *Journal of neuroscience methods*, *133*(1), 173-179.

[18] Pack, A. I., Galante, R. J., Maislin, G., Cater, J., Metaxas, D., Lu, S., ... & Peters, L. L. (2007). Novel method for high-throughput phenotyping of sleep in mice. *Physiological genomics*, *28*, 232.

[19] Fisher, S. P., Godinho, S. I., Pothecary, C. A., Hankins, M. W., Foster, R. G., & Peirson, S. N. (2012). Rapid assessment of sleep-wake behavior in mice. *Journal of Biological Rhythms*, *27*(1), 48-58.

[20] Martin Bland, J., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307-310.

Chapter 2

[1] Donohue, K. D., Medonza, D. C., Crane, E. R., & O'Hara, B. F. (2008). Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice. *Biomedical engineering online*, *7*(1), 14.

[2] Medonza, D.C. (2006). *Automatic detection of sleep and wake states in mice using Piezoelectric sesnors (Master's thesis)*. University of Kentucky, Lexington, KY.

[3] Friedman, L., Haines, A., Klann, K., Gallaugher, L., Salibra, L., Han, F., & Strohl, K. P. (2004). Ventilatory behavior during sleep among A/J and C57BL/6J mouse strains. *Journal of Applied Physiology*, *97*(5), 1787-1795.

[4] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems.*Annals of eugenics*, *7*(2), 179-188.

[5] Donohue, K. D., Huang, L., Burks, T., Forsberg, F., & Piccoli, C. W. (2001). Tissue classification with generalized spectrum parameters. *Ultrasound in medicine & biology*, *27*(11), 1505-1514.

Chapter 3

[1] Friedman, L., Haines, A., Klann, K., Gallaugher, L., Salibra, L., Han, F., & Strohl, K. P. (2004). Ventilatory behavior during sleep among A/J and C57BL/6J mouse strains. *Journal of Applied Physiology*, *97*(5), 1787-1795.

[2] Patwardhan, A., Moghe, S., Wang, K., Cruise, H., & Leonelli, F. (1998). Relation between ventricular fibrillation voltage and probability of defibrillation shocks: Analysis using Hilbert transforms. Journal of electrocardiology, 31(4), 317-325.

[3] Proakis, J., & Manolakis, D. (2007). *Digital Signal Processing: Principles, Algorithms, and Applications* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

[4] J. F. Kaiser "On a simple algorithm to calculate the 'energy' of a signal", Proc. IEEE International Conf. Acoust., Speech, and Signal Processing, 1990

[5] Teager, H. M., and Teager, S. M. (1985). "A phenomenological model for vowel production in the vocal tract," in Speech Science: Recent Advances, edited by R. G. Daniloff (College-Hill, San Diego), pp. 73–109.

[6] Teager, H. M., and Teager, S. (1990) "Evidence for nonlinear sound production mechanisms in the vocal tract," in Speech Production and Speech Modelling, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, London), pp. 241–262.

[7] Shannon, C. E. (1949). Communication in the presence of noise. Proceedings of the IRE, 37(1), 10-21.

Chapter 4

[1] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems.*Annals of eugenics*, *7*(2), 179-188.

[2] Donohue, K. D., Medonza, D. C., Crane, E. R., & O'Hara, B. F. (2008). Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice. *Biomedical engineering online*, *7*(1), 14.

VITA

Ryan Gooch was born in Somerset, Kentucky. He attended the University of Kentucky and obtained his Bachelor of Science degree in Electrical Engineering in 2010. Following completion of his undergraduate studies, Ryan worked as an electrical engineer for Mason & Hanger. In 2012 Ryan returned to the University of Kentucky to obtain his Master of Science in Electrical Engineering and will graduate in 2014. In addition to his studies, Ryan worked as a Research Assistant and Teaching Assistant for the University of Kentucky, and as a Signal Processing Engineer for Signal Solutions. He is pursuing the publication of two articles in peer-reviewed journals, based on the research in his master's thesis.