

**OPTIMIZING OF THE AUTOMATED EXTRACTION OF  
AUDIBLE MOUSE VOCALIZATIONS**

by  
Zoe L. Shteyn

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the  
requirements for the degree of Master of Science in Neuroscience

Summer 2024

Copyright 2024 Zoe Shteyn  
All Rights Reserved

**OPTIMIZING OF THE AUTOMATED EXTRACTION OF  
AUDIBLE MOUSE VOCALIZATIONS**

by  
Zoe L. Shteyn

Approved: \_\_\_\_\_  
Dr. Joshua P. Neunuebel, Ph.D.  
Professor in charge of thesis on behalf of the Advisory Committee

Approved: \_\_\_\_\_  
Dr. Robert West, Ph.D.  
Chair of the Department of Psychological and Brain Sciences

Approved: \_\_\_\_\_  
Debra H. Norris  
Interim Dean of the College of Arts and Sciences

Approved: \_\_\_\_\_  
Louis F. Rossi, Ph.D.  
Vice Provost for Graduate and Professional Education and  
Dean of the Graduate College

## TABLE OF CONTENTS

LIST OF FIGURES.....	v
ABSTRACT.....	vii
Chapter	
1	INTRODUCTION.....1
1.1	Background.....1
1.2	Significance.....4
1.3	Critical Barriers in the Field.....5
1.4	Rationale.....7
2	METHODS.....9
2.1	Experimental Subjects.....9
2.2	Experimental Setup.....9
2.3	Analyses.....12
2.4	Statistical Approaches.....14
2.5	Features to Optimize for Automatic Extraction.....16
3	RESULTS.....20
3.1	Curation of the Dataset.....20
3.2	Characterization of the Dataset.....21
3.3	Adjustment of the P-Value.....22
3.4	Adjustment of the Convolution Size.....22
3.5	Adjustment of the Minimum Object Area.....22
3.6	Adjustment of the Harmonic Parameters.....23
4	DISCUSSION AND CONCLUSION.....31
4.1	Summary of Results.....31
4.2	Limitations.....32
4.3	Future Directions.....33
4.4	Broad Applications.....34
REFERENCES.....	35

Appendix

A	IACUC Protocol Approval.....	41
B	Data Storage.....	42

## LIST OF FIGURES

- Figure 1** Experimental design showing two mice, one male and one female, interacting in a clean, empty cage within an anechoic chamber. The anechoic chamber has soundproofing foam along each wall to prevent outside noise from interfering with the recording. One microphone set 12.5 inches above the cage records vocalizations emitted during a 3-minute interaction.....11
- Figure 2** Labeled spectrogram displaying several vocalizations in time. Intensity of the pixel is directly related to intensity of sound, meaning louder sounds create more power on the spectrogram and therefore show up darker pixelated. Notation of the start time, end time, low frequency, and high frequency of each vocalization in a recording is done according to spectrotemporal structure. This data is then inputted into a table that clearly identifies characteristics of all audible vocalizations in a recording.....13
- Figure 3** Histograms of each measured characteristic, A) high frequency, B) low frequency, C) bandwidth, and D) duration are shown above. All histograms display a non-normal distribution of data. This is due to the lack of symmetry surrounding the mean and extreme values.....15
- Figure 4** Convolution size begins with a kernel that is used as the selected area of input. This selected area is a starting point for a more concise output area. Vocal signals may pass through the input area and can be identified better with an area that is slightly larger than the area of a vocal signal at a given point.....18
- Figure 5** The minimum object area can be visualized as the area of a vocalization segment on a frequency by time graph. More accurate analysis of the area of a vocal signal can be made by adjusting the frequency per second input.....19
- Figure 6** The number of audible vocalizations in each recording, as well as the total number of audible vocalizations across all recordings.....24

**Figure 7** Characterization of the dataset used four characteristics of interest. A) The high frequency distribution, B) low frequency distribution, C) bandwidth distribution, and D) duration distribution across all six recordings can be observed. Each characteristic contains a box plot of the distribution of data for each recording. For each box plot, the red line shows the median, or the 50th percentile of the data. The top and bottom of the box show the 75th and 25th percentile of the data, respectively. Dotted lines extending from the top and bottom of the box show variability outside the upper and lower quartiles. Extreme outliers are identified as + signs above or below the data distribution.....25

**Figure 8** The Kruskal Wallis statistical test code also corrected for multiple comparisons of means. This table shows the p-values of multiple comparisons run across all recordings to ensure that the probability of finding a false positive remained low. Highlighted boxes show where there was a statistically significant difference between recordings.....26

**Figure 9** Manipulation of the p-value from A) 0.05 to B) 0.01 displays a decrease in accuracy of extraction. This shows that a p-value of 0.05 will extract vocal signals with more accuracy.....27

**Figure 10** Adjustment of the convolution size from A) 1000, 0.001 to B) 500, 0.001 showed a decrease in extraction of vocal signals through decrease in frequency. The larger area of input provides better accuracy for the extraction process....28

**Figure 11** Modification of the minimum object area to a smaller value from A) 18.75 to B) 4.6875 improved extraction. A more concise area was needed for the program to accurately identify vocal signals.....29

**Figure 12** Cropped spectrogram of a recording showing A) harmonic overlap, ratio, and fraction set to value of 0.5 and B) adjustments made to harmonic overlap (value of 0.00001), ratio (0.99999), and fraction (0.00001) to increase optimization of extraction.....30

## **ABSTRACT**

Animals interact socially for the purpose of exchanging information, coordinating activities, and establishment of social relationships. These interactions can involve the emission of acoustic signals to facilitate communication between individuals. Mice have a large vocal repertoire with vocalizations emitted in the audible range of human hearing and extending above into the ultrasonic range. Mouse vocalizations of a higher frequency have been characterized extensively in efforts to understand how these acoustic signals influence behavior. However, knowledge of audible vocalizations, such as their various spectrotemporal structures, remains limited. Spectrotemporal structures refers to the time and frequency domains of a vocalization that provides a visual shape to observe. To overcome this gap in knowledge, first the manual curation and characterization of a dataset of audible vocalizations emitted from opposite-sex mouse social interactions was executed. Then, the adjustment of custom-written software that had been previously optimized for ultrasonic vocalization extraction was applied to dyadic social interactions that contained audible vocalizations. Comprehensive statistical tests were used to determine significant differences across recordings. The custom-written software was able to locate and extract parts of a whole vocalization within a recording, showing progress in our efforts to optimize automated extraction of audible mouse vocalizations.

# CHAPTER 1

## INTRODUCTION

### 1.1 BACKGROUND

Communication serves as a fundamental mechanism for the exchange of information, coordination of activities, and establishment of social relationships (Kaidanovich-Beilin et al., 2011). Throughout the animal kingdom, there are a wide variety of vocal repertoires that can be studied to further our collective understanding of communication. Vocalizations enable animals to convey complex messages related to mating, territory, hierarchy, and danger (Brennan & Kendrick, 2006). Primates have been known to vocalize differentially in association with certain affective or motivational states (Fischer, 2021). Specifically, adult sooty mangabeys emit predator-specific alarm calls which alert others of potential danger. In cetacean societies, vocal signaling is the primary mode of communication and consists of a diverse repertoire including whistles, creaks, squawks, variable rate click trains, and bangs (Luís et al., 2021). According to a review conducted on the vocal repertoire of zebra finches, a domesticated songbird, there is a wide range of information that can be gathered from acoustic communication in regard to contextual emission and the importance of spectral shape and pitch (Elie & Theunissen, 2015). For instance, loud calls that carry information about an individual's whereabouts can help in assisting songbird partners to reunite after losing visual contact (Vignal et al., 2008). Furthermore, alarm calls can be predator specific,

allowing nearby individuals to make necessary decisions for their ultimate survival (Evans et al., 1993). Thus, the social context in which vocalizations are emitted can be used to advance behavioral research in both animals and humans.

Mice are important subjects of behavioral research due to their diverse behaviors and vocalizations, as well as their ability to thrive in most environments (Neunuebel et al., 2015). In particular, mice are an ideal model system that have been used to advance the behavioral field through analysis of their auditory communication (Grimsley et al., 2016; Ihnat et al., 1995; Niemczura et al., 2020; Sangiamo et al., 2020). These studies have found that vocal signals are different among different behavioral contexts, revealing a link between vocalizations, behaviors, and affective states that requires a more in depth evaluation. Sounds vary in frequency; therefore, mouse vocalizations can be grouped into either audible, within the human hearing range, or ultrasonic, above the human hearing range, based on their frequency (Gillam, 2011). Mice have a diverse vocal repertoire that spans the audible range of human hearing, which stops at around 20kHz, and continues into the ultrasonic range of above 35kHz (Sangiamo et al., 2020). Acoustic signals can be understood in terms of their temporal and spectral properties, or their time and frequency domains, respectively. A time domain graph shows how a signal changes over time, while a frequency domain graph shows how a signal is distributed within different frequency bands over a range of frequencies. Spectrograms allow the visualization of both qualities, resulting in a graph that displays the frequencies on the y-axis and how they change over time on the x-axis. Through spectrograms, further analysis and categorization of spectrotemporal

structures is possible. Ultrasonic vocalizations (USVs) vary in spectrotemporal structure, specifically in their slope, with different environmental conditions, the gender of surrounding individuals, and developmental stage (Finton et al., 2017). Moreover, the emission and structure of vocal signals depends on behavioral context. In a study conducted on the relationship between mouse vocalizations and behavior, evidence was found that male mice in a non-aggressive role were likely to emit ultrasonic vocal signals that ascend in pitch, showing a positive slope (Sangiama et al., 2020). Male mice that exhibited aggressive behaviors were likely to emit ultrasonic signals descending in pitch. Therefore, it was shown that specific behavior-dependent dominant vocal signals emitted by male mice modulated the behavior of a social partner. This provides insight into the function of social ultrasonic vocalizations during complex interactions.

However, research on the accurate and detailed extraction of audible mouse vocalizations is limited. Early reports describe audible vocalizations in the context of pain cries and defensive behaviors (Grimsley et al., 2016; Ruat et al., 2022). One study that aimed to better understand how stress is expressed in vocalizations of mice, it was reiterated that males emit low frequency vocalizations when in pain, during handling, and while fighting. More importantly, it was shown that structurally similar low frequency vocalizations are emitted by female mice during mating and distress (Grimsley et al., 2016). This makes the study of audible low frequency vocalizations important to understanding both female and male mouse behaviors. While there has been extensive categorization of USVs based on their acoustic structure, audible vocalizations exhibit a diverse repertoire of structures. A study that aimed to categorize both ultrasonic and audible mouse vocalizations

based on their spectrotemporal structure, as well as link structures to behaviors, concluded that there was a large variation in audible signals (Finton et al., 2017). This may likely be due to the presence of harmonics, which can start near 500 Hz and extend upwards of 100,000 Hz. Like USVs, audible vocalizations are emitted in multiple contexts, however in contrast to USVs, they contain a large proportion of spectral nonlinearities that can make identification of a whole audible vocalization more difficult.

## **1.2 SIGNIFICANCE**

Animals rely on a variety of sensory cues to perceive, gather, and share information about their environment, making sensory cues an important target when attempting to understand intraspecies communication. Analysis of individual behavioral outputs and the response of a receiver can provide information about the importance of sensory cues during a social interaction (Rogers et al., 2000). For instance, visual cues such as bobbing of the head and extension of a colorful throat fan of a male green anole signals territory ownership to other individuals in the area (Gillam, 2011). Similarly, vocalizations emitted during an interaction may reflect the internal state of both the sender and receiver (Niemiczura et al., 2020). Internal state can be defined as the dynamic between physiological and neuronal activity, consequently altering behavior. The internal state of an individual is determined by major life events like mating, parturition, and neonatal development, as well as short term emotions and hormonal levels (Sangiomo et al., 2020; Maney, 2013). Experimental interactions using lab animals may hold translational value on the analysis of social behavior in other species. Individual decision-making and overall group behavior is

influenced by a variety of sensory cues, creating an opportunity for researchers to better understand social dynamics through targeted sensory intervention (Gillam, 2011). Furthermore, increased knowledge about sensory systems can provide insight into neurodevelopmental disorders in which the ability to communicate is impaired (Balasco et al., 2019).

### **1.3 CRITICAL BARRIERS IN THE FIELD**

The ability to monitor dynamic social interaction within a group of animals has become more accessible through advancements in technology. This includes sensitive microphone systems that pick up both audible and ultrasonic sounds and custom-written computer software that controls and analyzes recordings in a time-efficient manner (Niemczura et al., 2020; Sangiamo et al., 2020; Warren et al., 2020; Ruat et al., 2022). Research on mouse USVs has evolved due to advancement of automatic extraction techniques leading to more detailed classification and analysis of behavior. Studies using playback of audible vocalizations to elicit anxiety-related behaviors show the importance of low frequency sound in communication between animals (Sangiamo et al., 2020). Through collection of vocal signals, deeper insight into contextualizing specific social states and behaviors is possible.

Mice tend to audibly vocalize more during interactions surrounding mating such as pursuit and rejection (Finton et al., 2017). However, few studies have examined audible vocalizations in the context of opposite-sex complex interactions. Furthermore, in most studies that have assessed vocal signals in the audible range, the method of extraction is

done manually by multiple observers (Finton et al., 2017; Grimsley et al., 2016). This lack of detailed knowledge in regard to contextual understanding and extraction techniques has created a gap in the field that automated programming could help to advance. Automation would allow for more efficient extraction of audible vocalizations and the opportunity to gain more information from complex social interactions.

Software programs, like HARKBird, have been optimized for birdsong vocalizations through a collection of Python scripts that aim to extract songs from multiple individuals automatically (Sumitani et al., 2021). HARKBird is an open-sourced robot audition software that can extract acoustic events in a recording, obtain start and end timings, spatial information derived from the position or direction from the microphone array, and spectrograms of the sound separated from the original recording. Another software that has been used to extract bird vocalizations is Chipper, which aims to facilitate syllable segmentation and analysis of frequency, duration, and syntax (Searfoss et al., 2020). However, Chipper is primarily designed to parse syllables from a short birdsong bout of around 0.5-10 seconds, limiting the use of the software on longer recordings. Mouse USVs have been thoroughly characterized in terms of their spectrotemporal structure and context in which they are produced through optimized audio extraction and segmentation software. In a study conducted on USVs emitted by females during social interaction, vocal signals were detected and extracted using custom-written software based on multi-taper spectral analysis (Ax, available at <https://github.com/JaneliaSciComp/Ax>), which aims to find tones that are significantly above background noise and produce bounding boxes consisting of start and stop times, and low and high frequencies (Warren et al., 2020). One study that

addressed the extraction of audible vocalizations in mice that were bred for high anxiety related-behavior used software called Raven Pro (available at <https://www.ravensoundsoftware.com/>) to analyze the number of calls and call duration manually (Ruat et al., 2022). Another study that observed both USVs and audible vocalizations used Avisoft Bioacoustics SAS lab software (available at <https://avisoft.com/sound-analysis/>) to automatically extract syllable duration, peak frequency, and the number of instances of vocalization type (Niemczura et al., 2020). Altogether, there is a wide range of softwares that are useful in the extraction of both ultrasonic and audible animal vocalizations. Knowledge about these softwares is crucial to the improvement of extraction techniques for the ultimate advancement of the behavioral field. However, while there are many different software programs to extract vocalizations from recordings of various animals, the parameters for audible mouse vocalizations have not been adjusted for accurate extraction. The standardization of the extraction process will further future objectives of characterizing audible vocalizations and understanding how sensory information modulates behavior and, ultimately, neural activity.

#### **1.4 RATIONALE**

Many studies have observed mouse vocalizations and corresponding actions in an effort to characterize specific social behaviors (Ihnat et al., 1995; Niemczura et al., 2015; Neunuebel et al., 2015; Sangiamo et al., 2020). Social interaction is a critical component for the survival of various species as these interactions can help build and maintain complex societies (Gillam, 2011). In mice, relationships with conspecifics are important for social

hierarchy and mating (Kaidanovich-Beilin et al., 2011). Mouse USVs have been widely researched due to their prevalence, consistent spectrotemporal structures, and defined frequency range (Yao et al., 2023). However, only a few studies have looked at audible vocalizations of mice in social settings, and even less have investigated the behavioral considerations. Studies have shown that many characteristics such as frequency range differ between audible and ultrasonic vocalization repertoire (Grimsley et al., 2016). It is important for the advancement of knowledge to standardize and optimize extraction software parameters that reflect characteristics of audible vocalizations. Therefore, the objective of Aim 1 is to first manually curate and characterize a rich data set of audible mouse vocalizations recorded during dyadic social interactions between males and females. It is crucial to create this dataset for the purpose of comparison when evaluating the accuracy of the extraction software. Next, I will attempt to optimize the automated software that will extract audible vocalizations for future analysis of behavior in mice. These are critical steps in forming the foundation for understanding the functional role audible vocalizations play in mouse social behavior. My approach for achieving this objective involves using dyadic social paradigms to elicit audible vocalizations. I will then refine custom-written software to optimize the automatic extraction of audible vocalizations for the purpose of future characterization and deciphering the role these sounds play in behavior.

## **CHAPTER 2**

### **METHODS**

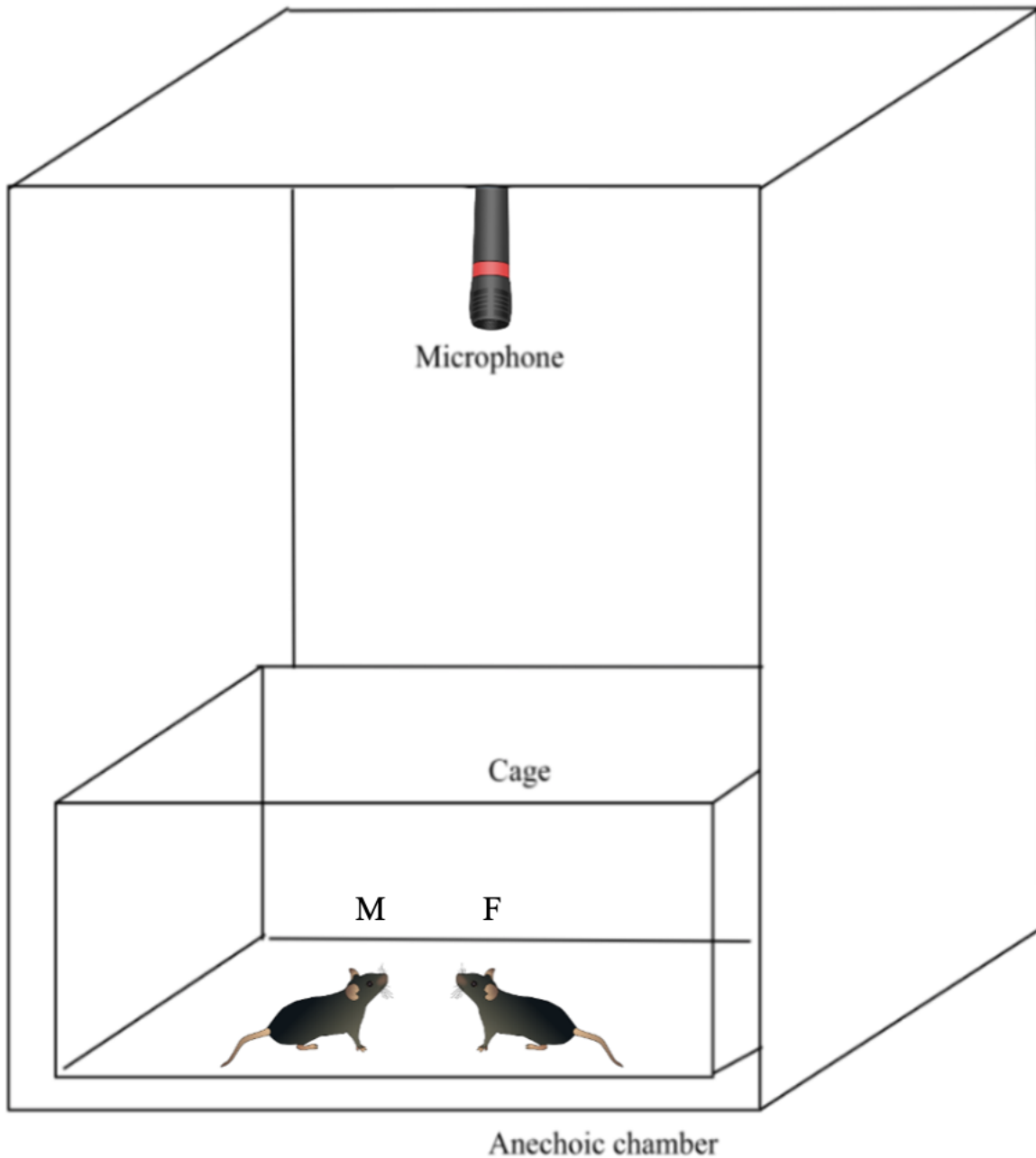
#### **2.1 EXPERIMENTAL SUBJECTS**

The experimental subjects were mice (n = 5 male; n = 5 female; aged 8-12 weeks) of the C57BL/6J strain which arrived at our facility at 8 weeks of age and were habituated to the animal colony. They were individually housed in cages containing ALPHA-dri bedding, environmental enrichment, and animals were allowed *ad lib* access to food and water. Mice were implanted with a light-activated microtransponder for identification. The colony room was kept on a 12/12 dark/light cycle (lights on at 9pm). All experiments were conducted during the dark phase. Experiments were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the University of Delaware Animal Care and Use Committee (protocol number: 1275-2024-0).

#### **2.2 EXPERIMENTAL SETUP**

Mice were recorded in an opposite-sex social interaction paradigm. There was a total of 6 recordings, which included one male and one female wild-type mouse that were 8-12 weeks of age. The anechoic chamber in which the experiments took place aimed to reduce background noise from interfering with the recording of vocalizations. Both mice were placed in a clean, sterile cage inside the chamber for a 3-minute period in which a single microphone

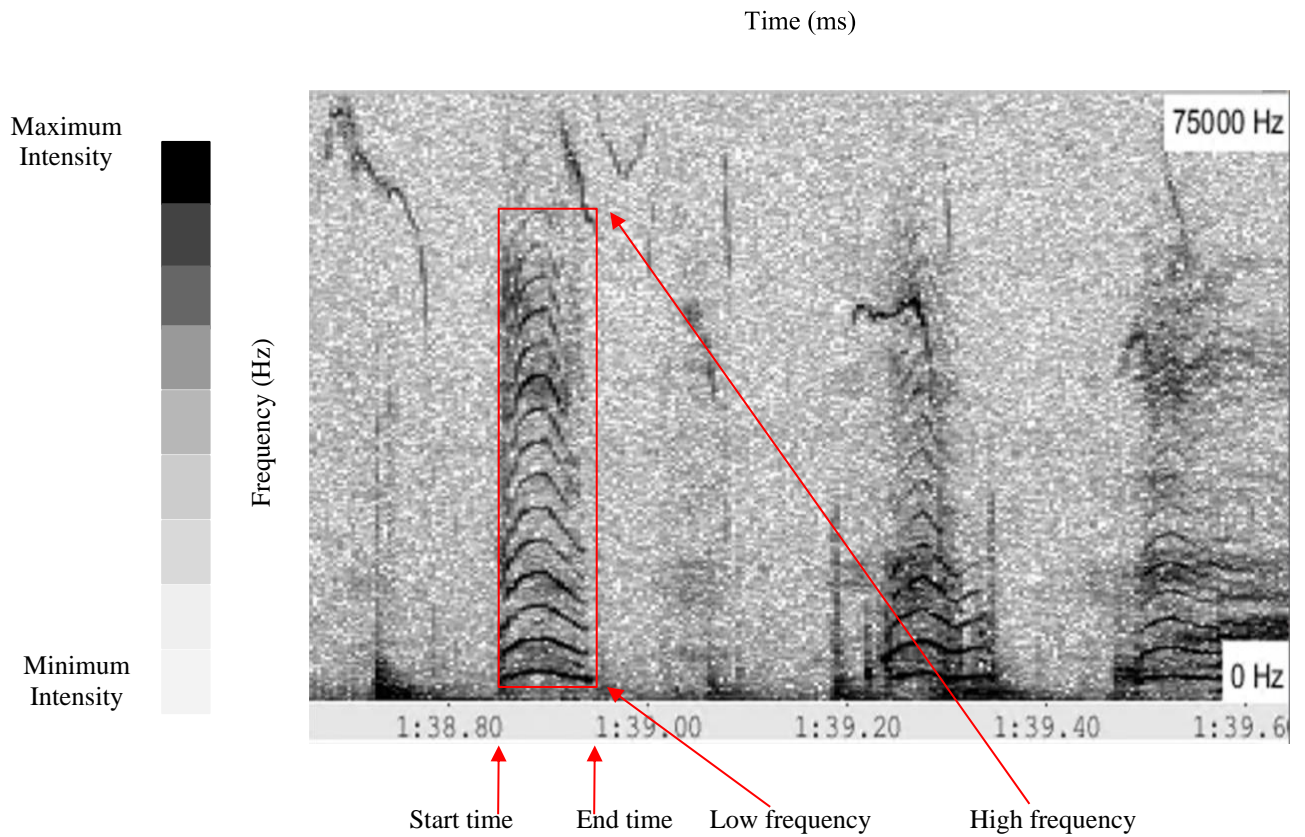
(Bruel & Kjaer, part number 4939), set 12.5 inches above the floor of the cage, recorded vocalizations at a sampling rate of 250,000 frames per second. Multiple recordings were made in one day and each time a new clean, sterile cage was used to maintain a consistent and controlled environment. The audio data was amplified with a preamplifier (Bruel & Kjaer, part number 2670) and a second amplifier (Bruel & Kjaer, part number 2690 A 0S4). We used hardware from National Instruments (National Instruments; Austin, TX; PXIe-1073, PXIe-6356, BNC-2110) and low-pass filtered at 200 kHz (Krohn-Hite, Brockton, MA; Model 3384) to record the audio data. Custom-written Matlab software (Mathworks; Natick, MA; version 2014b), was used to control all recording devices. All audio and video data were stored on a PC (Hewlett-Packard; Palo Alto, CA; Z620). Figure 1 depicts the experimental design for the experimental dyadic recordings.



**Figure 1.** Experimental design showing two mice, one male and one female, interacting in a clean, empty cage within an anechoic chamber. The anechoic chamber has soundproofing foam along each wall to prevent outside noise from interfering with the recording. One microphone set 12.5 inches above the cage records vocalizations emitted during a 3-minute interaction.

## 2.3 ANALYSES

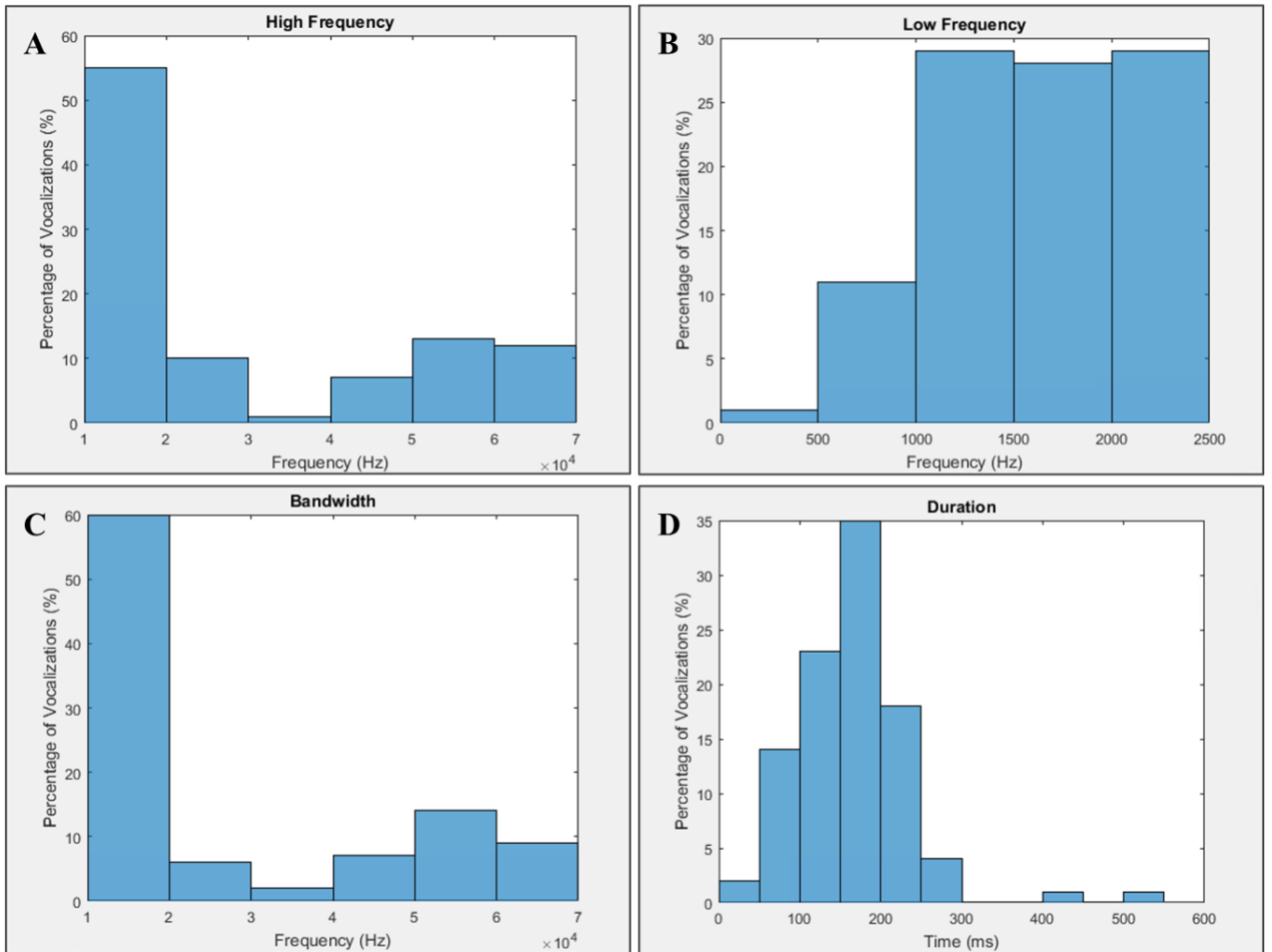
The manual scoring of vocalizations provides an accurate dataset for comparison to the output of extraction of the custom-written computer software. Tempo, a Matlab graphical user interface (available at <https://github.com/JaneliaSciComp/tempo/blob/master/Tempo.m>) is used for browsing, analyzing and annotating time-synchronized media files. This software produces a spectrogram per 3-minute recording showing the spectrotemporal structure of emitted vocalizations which are then manually analyzed for their start and stop time, high and low frequency, bandwidth, and duration. Intensity of a pixel on the spectrogram is directly related to the intensity of sound being produced. Figure 2 shows a labeled spectrogram from which specific data can be collected.



**Figure 2.** Labeled spectrogram displaying several vocalizations in time. Intensity of the pixel is directly related to intensity of sound, meaning louder sounds create more power on the spectrogram and therefore show up darker pixelated. Notation of the start time, end time, low frequency, and high frequency of each vocalization in a recording is done according to spectrotemporal structure. This data is then inputted into a table that clearly identifies characteristics of all audible vocalizations in a recording.

## 2.4 STATISTICAL APPROACHES

A statistical test must be used to determine if there are significant differences between recordings. Each measured characteristic will be tested for normality through observation of the histograms. Figure 3 portrays histograms for the four measured characteristics of the data which are high frequency, low frequency, bandwidth, and duration. As can be observed, the distribution of data in all histograms is non-normal and therefore requires a statistical test that accounts for this distribution. The Kruskal-Wallis test is a non-parametric statistical test used to compare two or more groups of an independent variable on a continuous dependent variable. This test assumes that all samples come from populations having the same continuous distribution, meaning it should provide information on if there are any significant differences between recordings. We are using this test because it does not require the assumptions of normality that other parametric tests do and can be used with small sample sizes. Furthermore, the Dunn-Šidák correction method adjusts the significance level when conducting multiple comparisons of means to control the family-wise error rate. This method ensures that the overall probability of making one or more Type I errors remains within the desired threshold. Altogether, the statistical tests will be used to identify if there are significant differences in the characteristics of audible vocalizations across recordings. This will be helpful in understanding how to better optimize parameters for extraction.



**Figure 3.** Histograms of each measured characteristic, A) high frequency, B) low frequency, C) bandwidth, and D) duration are shown above. All histograms display a non-normal distribution of data. This is due to the lack of symmetry surrounding the mean and extreme values.

## 2.5 FEATURES TO OPTIMIZE FOR AUTOMATIC EXTRACTION

Curating and characterizing a data set of audible vocalizations was an intensive, time-consuming stage of the project, which is a critical step necessary for assessing the automated extraction of audible vocalizations. Next, we attempted to extract audible sound using our custom-written software by adjusting multiple parameters to optimize the segmentation process (Neunuebel et al., 2015). The parameters we focused on adjusting are the p-value, convolution size, minimum object area, merge harmonics overlap, merge harmonics ratio, and merge harmonics fraction.

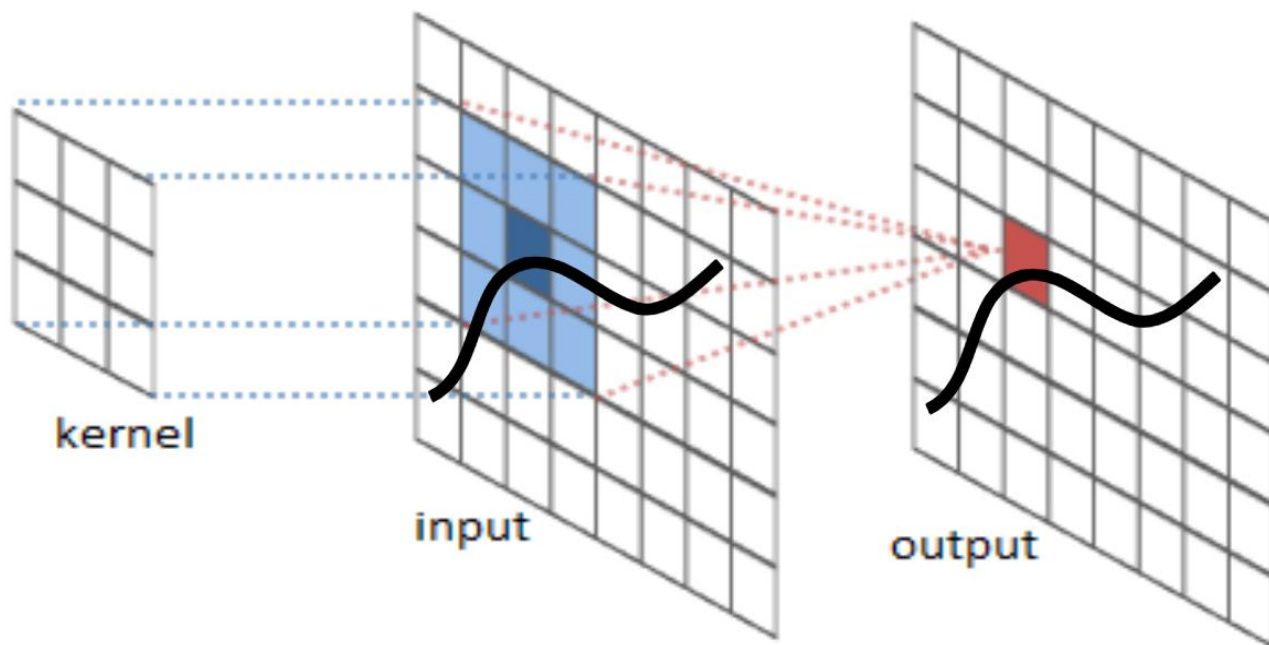
An F-test is used to determine if a pixel in a spectrogram is significantly above the background within a recording, and the significance level is conveyed with p-value. The pixels that are above the background are classified by our automated vocal extraction program as sounds (Neunuebel et al., 2015). If the p-value is less than 0.05, then the result is significant. Adjusting the p-value will affect extraction by changing the significance level of a vocal signal in comparison to the background. Increasing the threshold for significance (i.e., dropping alpha levels and p-values) enforces stricter criteria for inclusion as a sound.

Convolution size refers to enhancing the resolution of a small target area within a larger area using a grid system. The grid is created on top of the spectrogram through input of time in seconds by frequency in Hz. A diagram depicting convolution size can be found in Figure 4. Adjusting convolution size will affect extraction by changing the accuracy of vocal signal recognition within a target area.

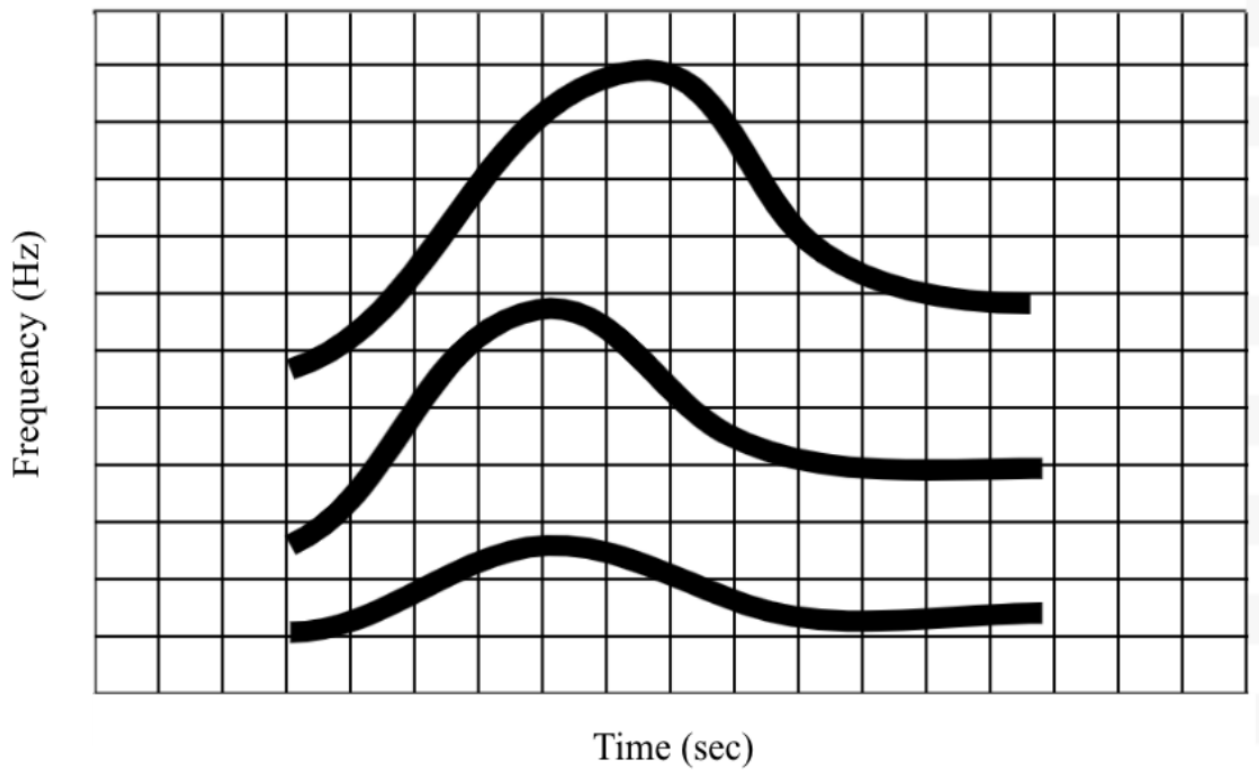
Minimum object area is a parameter for the system to distinguish a vocalization segment from background noise. Similar to convolution size, the region that is being

measured is made of frequency and time. However, the minimum object area is calculated using frequency in Hz per 1 second, which will provide insight into the average area of a vocalization segment. This can be observed in Figure 5. Minimum object area will affect extraction by providing a more accurate analysis of the area of a vocal signal for the system to recognize.

Harmonics can be defined as multiple instantaneous frequencies appearing in a vertical stepwise manner within the same time period (Vogel et al., 2019). The harmonic parameters will need to be adjusted due to the presence of harmonics in the majority of audible vocalizations. Specifically, the harmonic parameter of importance is the ratio, which refers to the tolerance in frequency ratio that two segments must be within. This means that the system should be able to recognize harmonics that have slightly different frequencies but that are within the same timeframe. The different amplitudes are multiples of the fundamental frequency, producing one harmonic. The other harmonic parameters that will be refined are the harmonic overlap, which determines the fraction in time two segments must overlap, and the harmonic fraction, which addresses the fraction of the overlap that must be within the ratio tolerance. Adjusting harmonic parameters will affect extraction due to their prevalence in the majority of audible vocalizations. This will assist the system in identifying audible vocalizations as well as their entire harmonic rather than just the fundamental frequency.



**Figure 4.** Convolution size begins with a kernel that is used as the selected area of input. This selected area is a starting point for a more concise output area. Vocal signals may pass through the input area and can be identified better with an area that is slightly larger than the area of a vocal signal at a given point.



**Figure 5.** *The minimum object area can be visualized as the area of a vocalization segment on a frequency by time graph. More accurate analysis of the area of a vocal signal can be made by adjusting the frequency per second input.*

## **CHAPTER 3**

### **RESULTS**

#### **3.1 CURATION OF THE DATASET**

There were 6 datasets of different dyadic pairings, all of which were manually extracted. To create a complete dataset, manual scoring of the raw data from spectrograms produced per 3-minute recording had to be meticulously curated. Audible mouse vocalizations contain a fundamental frequency, defined as the lowest and loudest resonant frequency (Nave, 2000). The fundamental frequency can be used to identify the time frame of a whole vocalization. Each audible vocalization in a recording has a start time, end time, low frequency, and high frequency that can be extrapolated from the spectrogram. The high frequency is determined by the highest harmonic present in a single vocalization. Harmonic overtones are integer multiples of the fundamental frequency (Green et al., 2019). Further data such as the total number, duration, and bandwidth of each vocalization in a recording was also noted. The duration can be calculated by subtracting the start time from the end time. Similarly, the bandwidth can be calculated by subtracting the low frequency from the high frequency.

### 3.2 CHARACTERIZATION OF THE DATASET

The number of vocalizations differed for each recording, ranging from 4 to 31, with an overall total of 98 vocalizations across all recordings. This can be seen clearly in figure 6. Similar to prior work (Warren et al., 2018), characteristics of the vocalizations such as duration, bandwidth, low frequency, and high frequency were calculated. Across all recordings, the low frequencies ranged from 110 Hz to 2,500 Hz while the high frequency had a much larger range of 12,000 Hz to 70,000 Hz. This can be seen in figure 7A and 7B, respectively. Therefore bandwidth, shown in figure 7C, was variable and had a range of 10,000 Hz to 67,000 Hz. Figure 7D shows that the durations of vocalizations ranged from 30 ms to 520 ms. When comparing the dyadic recording pairs across characteristics, there were a few significant differences. As identified in Figure 3, the distribution of the data for all characteristics was non-normal. Thus, the Kruskal Wallis test was used to determine if there were statistically significant differences between groups. Duration was significantly different only between recordings 1 and 5. Low frequency was significantly different for recordings 1 and 4. High frequency was significantly different for recordings 1 and 4, 1 and 5, 2 and 4, and 2 and 5. Lastly, bandwidth had similar results with significant differences between recordings 1 and 4, 1 and 5, 2 and 4, and 2 and 5. The Kruskal Wallis code also corrected for multiple comparisons of means. The Dunn–Šidák correction method adjusts the significance level when conducting multiple comparisons of means to control the family-wise error rate. This method ensures that the overall probability of making one or more Type I errors remains within the desired threshold. Figure 8 provides p-values for all recordings compared to each other. The significant differences seen across recordings will

influence the optimization of extraction, making it more difficult to adjust parameters to account for individual differences.

### **3.3 ADJUSTMENT OF THE P-VALUE**

Adjustment of the p-value from 0.05 to 0.01 proved to be detrimental to the optimization of vocalization extraction. As shown in figure 9, increasing the threshold for significance by decreasing the p-value enforced stricter criteria for inclusion of a sound, leading to less instances of extraction. Maintaining the p-value at 0.05 ensured statistical significance while still allowing for more optimal extraction of vocalizations.

### **3.4 ADJUSTMENT OF THE CONVOLUTION SIZE**

The modification of the frequency or time variables of the convolution size parameter changes the resolution of the area of extraction. By decreasing the frequency variable of the convolution size, as shown in figure 10, extraction was decreased. This is because it is more ideal for extraction of various vocalizations to have a larger range of tolerance in the extrapolated area. Decreasing the time variable would produce the same decreased extraction due to a smaller area of input.

### **3.5 ADJUSTMENT OF THE MINIMUM OBJECT AREA**

Reduction of the minimum object area allowed for better extraction. The difference in extraction can be seen in figure 11, whereby decreasing the frequency per 1 second created a more precise area from which to recognize harmonic segments of a vocalization. As the

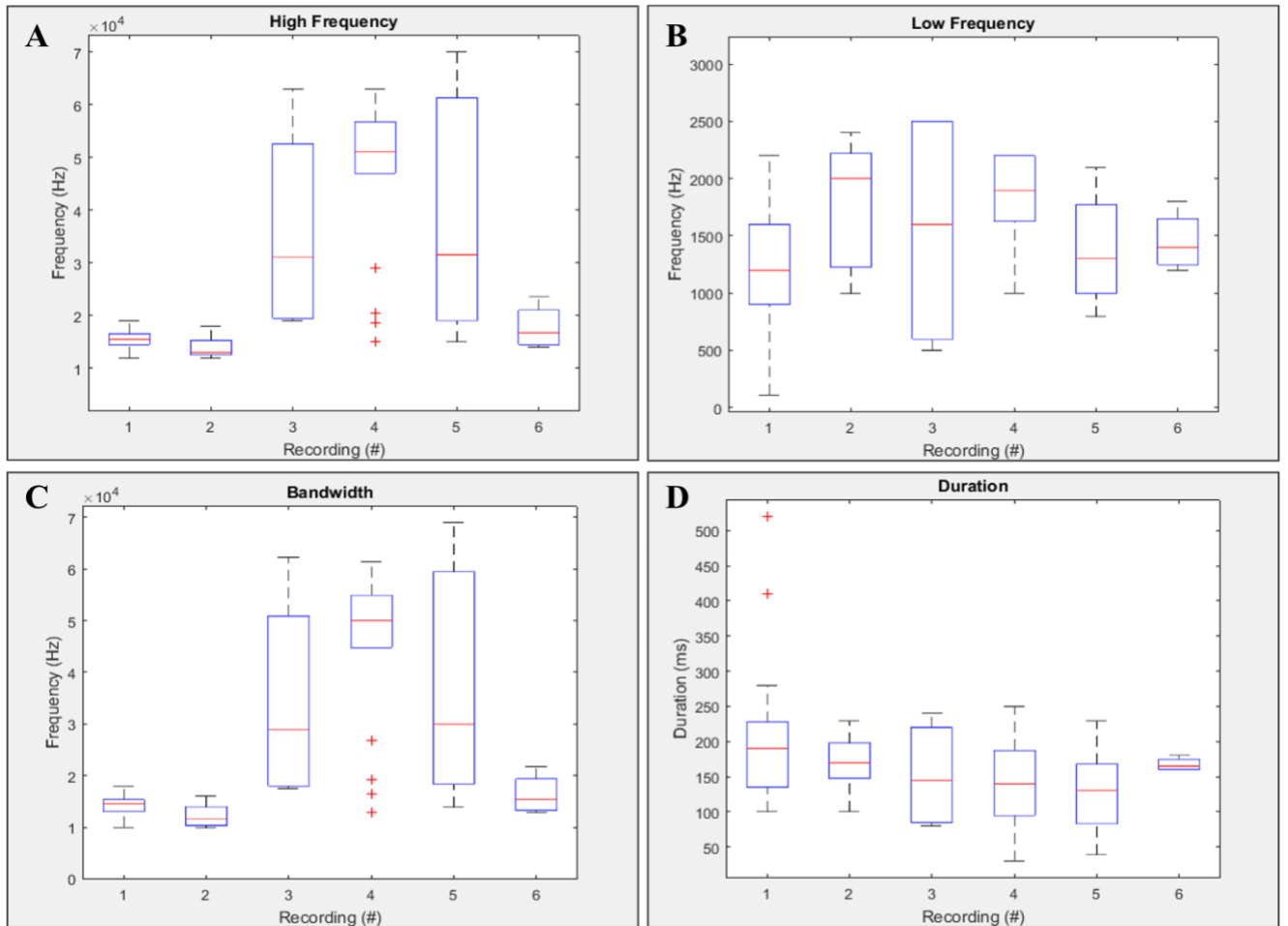
harmonics increase in frequency, they tend to decrease in power. This is translated onto the spectrogram as decreased area, therefore decreasing the minimum object area increases extraction output.

### **3.6 ADJUSTMENT OF THE HARMONIC PARAMETERS**

The parameters that produced the most change in extraction were the harmonic parameters. This can be seen in figure 12 which shows the difference in extraction between values of harmonic overlap, ratio, and fraction at 0.5 and 0.00001, 0.99999, and 0.00001, respectively. Harmonic overlap maintained a value of 0.00001 due to the need for the automated vocalization segmentation program to have a large amount of tolerance, or a smaller fraction, to capture harmonic segments that varied slightly in time. The harmonic ratio required a large tolerance as well to allow for variability in frequency between segments of one harmonic vocalization and was therefore set to a value of 0.99999. Lastly, the harmonic fraction also needed a small fraction value of 0.00001 to adjust for the variability in frequency and time of harmonic segments of a vocalization.

<b>Recording #</b>	<b>Vocalization #</b>
1	27
2	13
3	4
4	19
5	31
6	4
<b>Total:</b>	<b>98</b>

*Figure 6. The number of audible vocalizations in each recording, as well as the total number of audible vocalizations across all recordings.*



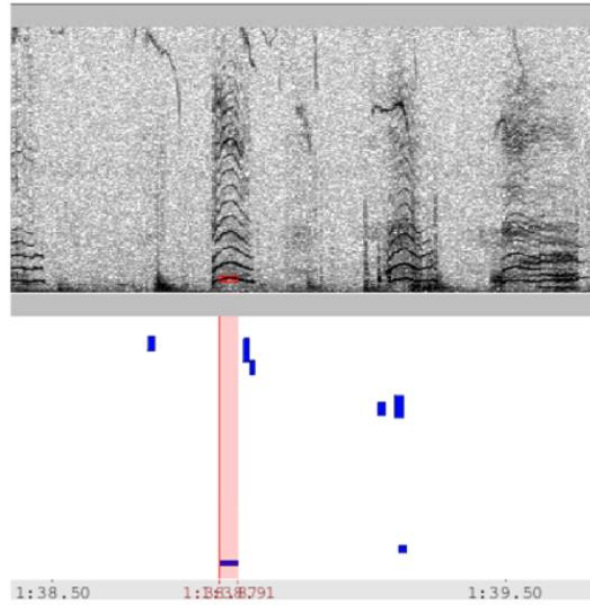
**Figure 7.** Characterization of the dataset used four characteristics of interest. A) The high frequency distribution, B) low frequency distribution, C) bandwidth distribution, and D) duration distribution across all six recordings can be observed. Each characteristic contains a box plot of the distribution of data for each recording. For each box plot, the red line shows the median, or the 50<sup>th</sup> percentile of the data. The top and bottom of the box show the 75<sup>th</sup> and 25<sup>th</sup> percentile of the data, respectively. Dotted lines extending from the top and bottom of the box show variability outside the upper and lower quartiles. Extreme outliers are identified as + signs above or below the data distribution.

Recording #		High Freq	Low Freq	Bandwidth	Duration
1	2	0.9976	0.0292	0.9827	1.0000
1	3	0.0903	0.9963	0.1090	0.9980
1	4	5.45E-07	0.0009	1.69E-06	0.2458
1	5	1.80E-07	0.9940	3.00E-07	0.0063
1	6	1.0000	1.0000	1.0000	1.0000
2	3	0.0256	0.9992	0.0207	1.0000
2	4	5.95E-07	1.0000	5.00E-07	0.9246
2	5	4.76E-07	0.2675	2.03E-07	0.3340
2	6	0.9763	0.9829	0.9528	1.0000
3	4	1.0000	0.9794	1.0000	1.0000
3	5	1.0000	1.0000	1.0000	0.9998
3	6	0.8599	1.0000	0.8778	1.0000
4	5	1.0000	0.0220	1.0000	0.9998
4	6	0.2649	0.8735	0.3215	0.9993
5	6	0.4030	1.0000	0.4183	0.9547

*Figure 8. The Kruskal Wallis statistical test code also corrected for multiple comparisons of means. This table shows the p-values of multiple comparisons run across all recordings to ensure that the probability of finding a false positive remained low. Highlighted boxes show where there was a statistically significant difference between recordings.*

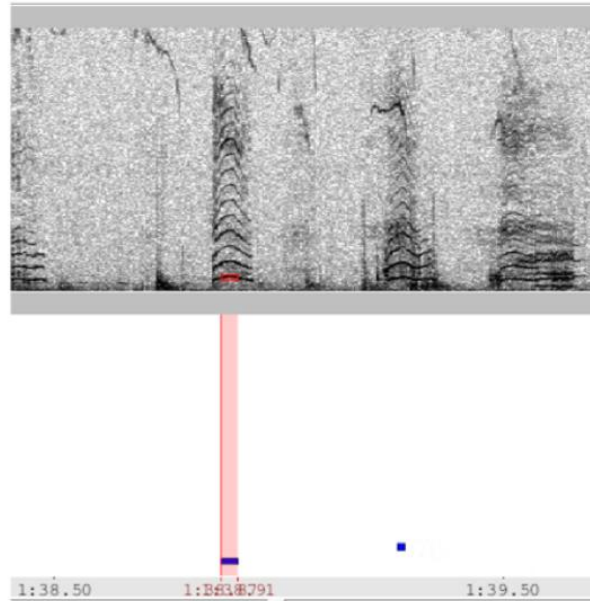
**A**

PVAL = 0.05



**B**

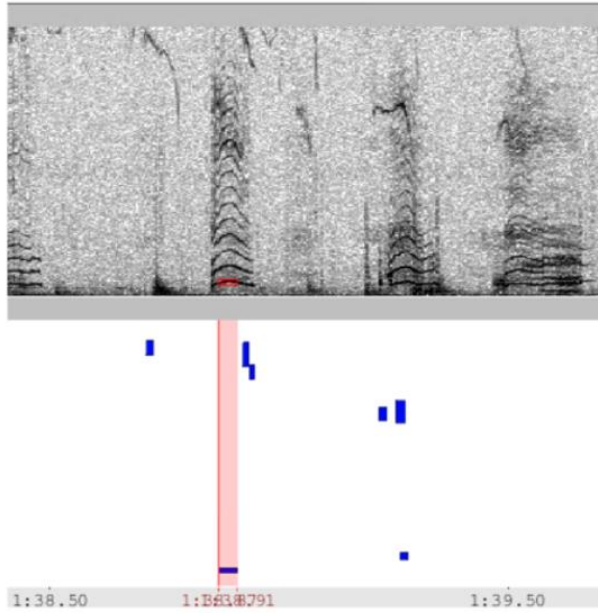
PVAL = 0.01



*Figure 9. Manipulation of the p-value from A) 0.05 to B) 0.01 displays a decrease in accuracy of extraction. This shows that a p-value of 0.05 will extract vocal signals with more accuracy.*

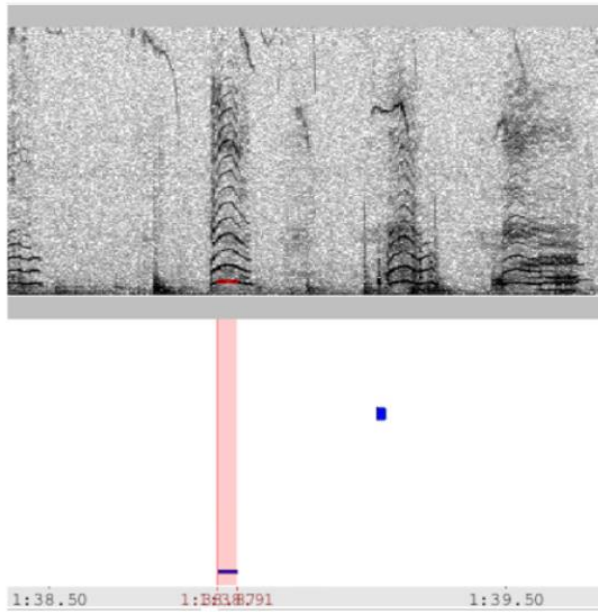
**A**

Convolution  
size =  
[1001,0.001]



**B**

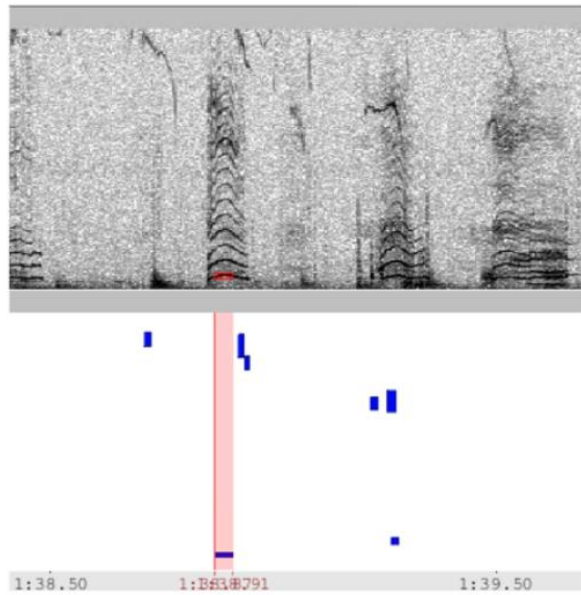
Convolution  
size =  
[500,0.001]



**Figure 10.** Adjustment of the convolution size from A) 1000, 0.001 to B) 500, 0.001 showed a decrease in extraction of vocal signals through decrease in frequency. The larger area of input provides better accuracy for the extraction process.

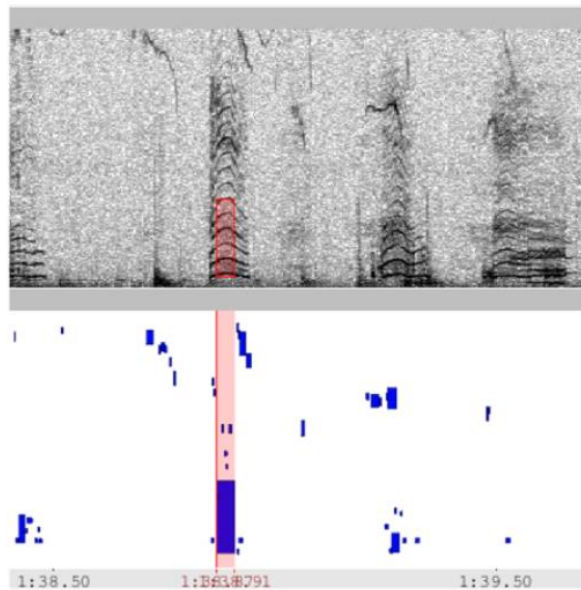
**A**

MOA = 18.75

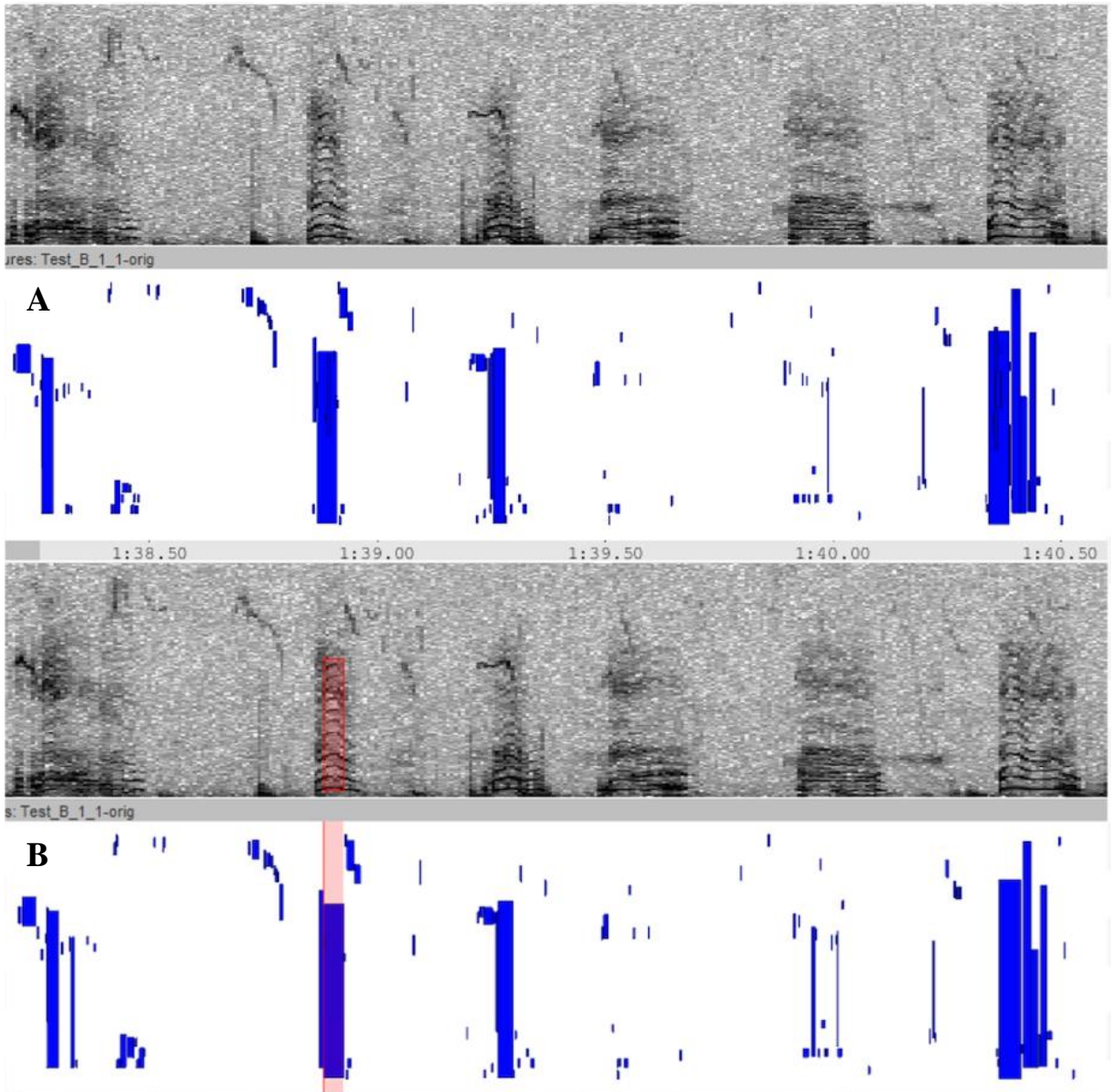


**B**

MOA = 4.6875



**Figure 11.** Modification of the minimum object area to a smaller value from A) 18.75 to B) 4.6875 improved extraction. A more concise area was needed for the program to accurately identify vocal signals.



**Figure 12.** Cropped spectrogram of a recording showing A) harmonic overlap, ratio, and fraction set to a value of 0.5 and B) adjustments made to harmonic overlap (value of 0.00001), ratio (0.99999), and fraction (0.00001) to increase optimization of extraction.

## **CHAPTER 4**

### **DISCUSSION AND CONCLUSION**

#### **4.1 SUMMARY OF RESULTS**

Refinement of our custom-written software to better extract audible vocalizations provided valuable insights into the quality of our optimization techniques. In figure 9, it can be observed that adjustment of the p-value parameter from 0.05 to 0.01 showed that a value of 0.05 allowed for better extraction. The value of 0.05 allowed for inclusion of audio signals while still maintaining statistical significance. Next, in figure 10, the manipulation of the convolution size showed better results when allowing for a larger range of tolerance. To extract a vocalization segment, the frequency and time variables had to address a large enough area of interest. Figure 11 demonstrated that the minimum object area had to be adjusted to account for a full vocal signal. Decreasing the frequency per second ratio increased extraction as the program was able to identify more harmonic frequencies of a vocalization. Additionally, we found that most audible vocalizations emitted during opposite-sex interactions consisted of harmonics extending into the ultrasonic range. This was important when considering manipulation of the harmonic parameters, as seen in figure 12. Harmonic overlap required a large amount of tolerance to account for varied overlap in the timing of harmonic frequencies, resulting in using a small value of 0.00001 to better extraction. This allowed there to still be recognition of harmonics as part of one vocalization

even when differing in timing. Similarly, harmonic ratio required a large tolerance to allow extraction of a vocalization with harmonic frequencies that varied in spectrotemporal structure. A value of 0.99999 for the harmonic ratio helped to improve extraction due to its high tolerance in frequency ratio that two segments must be within. Finally, the harmonic fraction parameter, which accounts for both the overlap and ratio parameters, needed a small value to allow for variability in the time and frequency of harmonics of a vocalization. Therefore, a value of 0.00001 led to improved extraction. Statistically significant differences between recordings were observed in a few instances. Duration was statistically different between recordings 1 and 5. Low frequency was statistically different between recordings 1 and 4. High frequency was significantly different for recordings 1 and 4, 1 and 5, 2 and 4, and 2 and 5. Lastly, bandwidth had significant differences between recordings 1 and 4, 1 and 5, 2 and 4, and 2 and 5. All results were corrected for multiple comparisons. These differences in recordings revealed a crucial aspect to consider during the optimization process. Optimization of extraction across recordings is more difficult due to individual differences and requires consideration when adjusting parameters. Moreover, performance of the adjustments made to parameters cannot be quantified at this time.

## **4.2 LIMITATIONS**

Our custom-written software was able to detect harmonic vocalizations, however the harmonics were extracted in parts. Merging of the parts is important to optimizing the extraction of a whole vocalization. Adjustments to the custom-written software, and more specifically the harmonic parameters, are critical to achieving optimal extraction within a

recording. The harmonic parameters would need more tolerance to allow for more extraction. Moreover, parameters such as the convolution size and minimum object may also require adjustment for more tolerance. Due to the ongoing nature of this project, other softwares may need to be considered for better extraction. These include Raven Pro and Avisoft Bioacoustics SAS lab software as well as softwares that have been optimized primarily for birdsong such as Chipper and HARKBird. Birdsong-centric softwares may offer better extraction due to the noted presence of harmonics in bird vocal repertoire (Elie & Theunissen, 2015). Moreover, importance should be placed on optimizing software for the recognition of the fundamental frequency in a harmonic audible vocalization. This may increase the rate of extraction as seen in studies that aim to extract based on the fundamental frequencies of ultrasonic vocal signals (Sangiromo et al., 2020).

#### **4.3 FUTURE DIRECTIONS**

Future directives surrounding more complex interactions will require optimization of an automated extraction software. Experimental interactions involving more mice or longer recordings emphasize the importance of accurately optimized automated extraction due to a large increase in data to analyze. Automatic extraction would provide more efficiency and less limitations to experimental designs that aim to understand elaborate behaviors. The consideration of other extraction software is strongly recommended, as the current system may not be optimizable for the extraction of entire audible vocalizations. Furthermore, combined with sound source localization software, the ability to identify the vocalizing individual and contextualize their behaviors, as well as other individuals in the social

paradigm, would be much more feasible. Similarly, optimized automated extraction would assist with naturalistic observation, creating a unique opportunity to study mouse audible vocalizations in an environment where there are fewer confounding variables.

#### **4.4 BROAD APPLICATIONS**

Mice are used as subjects in behavioral research due to their widely studied and extensive behavioral repertoire that can be used to inform human studies. An interesting application of the knowledge gained from audible mouse vocalizations would be to the similarities seen in qualities of human speech. For instance, a study showed that human speech sounds of vowels are made up of regularly spaced harmonics that are not of equal intensity (Schnupp et al., 2012). Regions of frequency space where sounds carry a lot of energy are called formants, which arise from resonance in the vocal tract. The resonance frequencies change with movement of the lips, jaws, tongue, and soft palate, which therefore changes the dimensions of the resonance cavities in the vocal tract. This produces audible vocalizations with harmonics that seem to jump up to a higher frequency, skipping frequencies in between. Although mice have distinctly different vocal tract morphology compared to humans, they can produce low frequency audible range vocalizations that display similarities in structure to human speech (Smith et al., 2020). As seen in the data collected from the dyadic mouse interactions, some audible vocalizations of mice have harmonics that skip frequencies, producing a variety of harmonic structures. This may be indicative of complex vocal patterns in mice as has been documented in human speech.

## REFERENCES

- Balasco, L., Provenzano, G., Bozzi, Y. "Sensory Abnormalities in Autism Spectrum Disorders: A Focus on the Tactile Domain, From Genetic Mouse Models to the Clinic." *Front Psychiatry*, 2019, doi: 10.3389/fpsy.2019.01016
- Brennan, P. A., & Kendrick, K. M. "Mammalian social odours: attraction and individual recognition." *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, pp. 2061-2078, 2006, doi:10.1098/rstb.2006.1931.
- Chung. "Diverse sensory cues for individual recognition." *Development, Growth & Differentiation*, 2020, doi:10.1111/dgd.12697.
- Delcourt, Matthieu, and Pablo A. Marquet. "The scaling of social interactions across animal species." *Scientific Reports*, vol. 11, 2021, article number 17856. doi: 10.1038/s41598-021-92025-1
- Fonseca, A.H.O., et al. "Analysis of ultrasonic vocalizations from mice using computer vision and machine learning" *eLife*, 2021, doi: 10.7554/eLife.59161
- Gillam, E. *An Introduction to Animal Communication*. Nature Education Knowledge 3(10):70, 2011
- Grimsley, J. M. S., et al. "Contextual Modulation of Vocal Behavior in Mouse: Newly Identified 12 kHz 'Mid-Frequency' Vocalization Emitted during Restraint." *Frontiers in Behavioral Neuroscience*, vol. 10, doi:10.3389/fnbeh.2016.00038.

Ihnat, R. et al. "Pup's broadband vocalizations and maternal behavior in the rat" Behavioural Processes, Volume 33, Issue 3, 1995, Pages 257-271, doi:10.1016/0376-6357(94)00028-F

Kaidanovich-Beilin, O., Lipina, T., Vukobradovic, I., Roder, J., & Woodgett, J. R. "Assessment of social interaction behaviors." Journal of visualized experiments: JoVE, no. 48, 2473, 2011, doi:10.3791/2473.

Niemczura, A. C., et al. "Physiological and Behavioral Responses to Vocalization Playback in Mice." Frontiers in Behavioral Neuroscience, vol. 14, doi:10.3389/fnbeh.2020.00155.

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., & Egnor, S. E. (2015). "Female mice ultrasonically interact with males during courtship displays." eLife, 4, e06203. <https://doi.org/10.7554/eLife.06203>

Rogers, Lesley J., and Gisela T. Kaplan. "Songs, Roars, and Rituals: Communication in Birds, Mammals, and Other Animals." Harvard University Press, 2000. <http://ndl.ethernet.edu.et/bitstream/123456789/569/1/60.pdf.pdf>

Sangiamo, D. T., Warren, M. R., & Neunuebel, J. P. "Ultrasonic signals associated with different types of social behavior of mice." Nat Neurosci, vol. 23, pp. 411-422, 2020, doi:10.1038/s41593-020-0584-z.

Vogel, A.P., Tsanas, A. & Scattoni, M.L. Quantifying ultrasonic mouse vocalizations using acoustic analysis in a supervised statistical machine learning framework. Sci Rep 9, 8100 (2019). doi: 10.1038/s41598-019-44221-3

Warren, M. R., Sangiamo, D. T. & Neunuebel, J. P. High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. *Journal of neuroscience methods* 297, 44-60, doi:10.1016/j.jneumeth.2017.12.013 (2018).

Yao, K., Bergamasco, M., Scattoni, M. L., & Vogel, A. P. (2023). A review of ultrasonic vocalizations in mice and how they relate to human speech. *the Journal of the Acoustical Society of America/the Journal of the Acoustical Society of America*, 154(2), 650–660. <https://doi.org/10.1121/10.0020544>

Finton, C. J., Keesom, S. M., Hood, K. E., & Hurley, L. M. (2017). What’s in a squeak? Female vocal signals predict the sexual behaviour of male house mice during courtship. *Animal Behaviour*, 126, 163–175. <https://doi.org/10.1016/j.anbehav.2017.01.021>

Elie, J. E., & Theunissen, F. E. (2015). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, 19(2), 285–315. <https://doi.org/10.1007/s10071-015-0933-6>

Xie, S., Lu, J., Liu, J., Zhang, Y., Lv, D., Chen, X., & Zhao, Y. (2022). Multi-view features fusion for birdsong classification. *Ecological Informatics*, 72, 101893. <https://doi.org/10.1016/j.ecoinf.2022.101893>

Grimsley, J. M. S., Sheth, S., Vallabh, N., Grimsley, C. A., Bhattal, J., Latsko, M., Jasnow, A., & Wenstrup, J. J. (2016). Contextual Modulation of Vocal Behavior in Mouse: Newly Identified 12 kHz “Mid-Frequency” Vocalization Emitted during Restraint. *Frontiers in Behavioral Neuroscience*, 10. <https://doi.org/10.3389/fnbeh.2016.00038>

Warren, M.R., et al. “Ultrashort-range, high-frequency communication by female mice shapes social interactions” *Scientific Reports*, vol. 10, 2020, doi: 10.1038/s41598-020-59418-0

Searfoss, A. M., Pino, J. C., & Creanza, N. (2020). Chipper: Open-source software for semi-automated segmentation and analysis of birdsong and other natural sounds. *Methods in Ecology and Evolution*, 11(4), 524–531. <https://doi.org/10.1111/2041-210x.13368>

Sumitani, S., Suzuki, R., Arita, T., Nakadai, K., & Okuno, H. G. (2021). Non-Invasive Monitoring of the Spatio-Temporal Dynamics of Vocalizations among Songbirds in a Semi Free-Flight Environment Using Robot Audition Techniques. *Birds*, 2(2), 158–172. <https://doi.org/10.3390/birds2020012>

Ruat, J., Genewsky, A. J., Heinz, D. E., Kaltwasser, S. F., Canteras, N. S., Czisch, M., Chen, A., & Wotjak, C. T. (2022). Why do mice squeak? Toward a better understanding of defensive vocalization. *iScience*, 25(7), 104657. <https://doi.org/10.1016/j.isci.2022.104657>

Maney, D. L. (2013). The incentive salience of courtship vocalizations: Hormone-mediated ‘wanting’ in the auditory system. *Hearing Research*, 305, 19–30. <https://doi.org/10.1016/j.heares.2013.04.011>

Nave, R. (2000). Fundamental and harmonic resonances. (n.d.). <http://hyperphysics.phy-astr.gsu.edu/hbase/Waves/funhar.html>

Green, D. M., Scolman, T., Guthrie, O. W., & Pasch, B. (2019). A broad filter between call frequency and peripheral auditory sensitivity in northern grasshopper mice (*Onychomys leucogaster*). *Journal of Comparative Physiology. A, Sensory, Neural, and Behavioral Physiology/Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 205(4), 481–489. <https://doi.org/10.1007/s00359-019-01338-0>

Fischer, J. (2021). Primate vocal communication and the evolution of speech. *Current Directions in Psychological Science*, 30(1), 55–60.

<https://doi.org/10.1177/0963721420979580>

Luís, A. R., May-Collado, L. J., Rako-Gospic, N., Gridley, T., Papale, E., Azevedo, A., Silva, M. A., Buscaino, G., Herzing, D., & Santos, M. E. D. (2021). Vocal universals and geographic variations in the acoustic repertoire of the common bottlenose dolphin. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-90710-9>

Vignal C, Mathevon N, Mottin S. Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behav Process*. 2008;77:191–198.

Evans CS, Evans L, Marler P. On the meaning of alarm calls: Functional reference in an avian vocal system. *Anim Behav*. 1993;46:23–38.

Ehret, G., & Bernecker, C. (1986). Low-frequency sound communication by mouse pups (*Mus musculus*): wriggling calls release maternal behaviour. *Animal Behaviour*, 34(3), 821–830. [https://doi.org/10.1016/s0003-3472\(86\)80067-7](https://doi.org/10.1016/s0003-3472(86)80067-7)

Warren, M. R., Spurrier, M. S., Roth, E. D., & Neunuebel, J. P. (2018). Sex differences in vocal communication of freely interacting adult mice depend upon behavioral context. *PloS One*, 13(9), e0204527. <https://doi.org/10.1371/journal.pone.0204527>

Schnupp, J., Nelken, I., King, A.J. (2012). Formants and harmonics in spoken vowels | *Auditory Neuroscience*. (n.d.). <https://auditoryneuroscience.com/vocalizations-speech/formants-harmonics>


Smith, S. K., Burkhard, T. T., & Phelps, S. M. (2020). A comparative characterization of laryngeal anatomy in the singing mouse. *Journal of Anatomy*, 238(2), 308–320.  
<https://doi.org/10.1111/joa.13315>

## Appendix

### A. IACUC PROTOCOL APPROVAL

**University of Delaware  
Institutional Animal Care and Use Committee  
Application to Use Animals in Research  
(New and 3-Yr submission)**

<b>Title of Protocol: Identifying the effect of vocal communication on social interaction and examining the impact of social cues on neural coding.</b>	
<b>AUP Number: 1275-2023-0</b>	<b>← (4 digits only — if new, leave blank)</b>
<b>Principal Investigator: Joshua P. Neunuebel</b>	
<b>Common Name (Strain/Breed if Appropriate): Mouse</b> <b>Genus Species: <i>Mus musculus</i></b>	
<b>Date of Submission: 7/18/2023</b>	

<b>Official Use Only</b>
IACUC Approval Signature: 
Date of Approval: <u>10.1.2023</u>

## **B. DATA STORAGE**

Each recording was saved to the W drive in a folder titled 'Behavioral Recordings' under the date when it was recorded. On the Z drive, all data was stored under the folder 'Zoe' including statistical analyses, Matlab functions, and data collection. The recordings used in this document were organized into a spreadsheet titled 'Voc Manual Scoring' which can be found in the folder titled 'Data' on the Z drive. The Matlab function of the Kruskal-Wallis test (available at <https://www.mathworks.com/help/stats/kruskalwallis.html>) can be found in the folder titled 'Matlab Functions' on the Z drive. The spreadsheet for multiple comparisons can be found in the folder 'Data Collection' on the Z drive. Statistical analyses of the characteristics of the data can be found in the folder titled 'Statistical Analyses' on the Z drive.