Developing a Behavioral Assay for Tinnitus Characterization

WPI

UMass Chan MEDICAL SCHOOL

A Major Qualifying Project submitted to the faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the degree of Bachelor of Science

Submitted By:

Myah Caplan

Jacob Mills

Submitted to:

Professor Adam Lammert, PhD

Submitted on:

April 27, 2023

Abstract

Tinnitus– affecting ~50 million Americans– is hard to characterize because of its diverse manifestations, which hinder treatment efficacy. Our goal was to further develop a pre-clinical tinnitus characterization assay using reverse correlation, where patients render subjective perceptions from random stimuli. We evaluated stimulus generation methods: an area identified for refinement. The most accurate characterizations came from the Brimijoin Gaussian Smoothed method; 8 segments on a frequency spectrum are systematically filled with a Gaussian-shaped power distribution. This showed statistically significant improvement and had the most positive subjective feedback. In the future, this research may be incorporated into a clinical setting to improve tinnitus treatment via characterization.

Table of Contents

Executive Summary

Introduction and Literature Review

Tinnitus is a prevalent hearing condition affecting roughly 50 million Americans and 3%-30% of people worldwide (Kaylie, 2022; Tunkel, 2014; Sanchez, 2004). Patients report their tinnitus sounding like various sounds, pitches, and volumes. The psychoacoustic tinnitus spectrum (PTS) refers to these sounds. It often ranges in severity but can lead to secondary symptoms such as insomnia, frustration, depression, and anxiety. While there are several available sound therapy techniques– including cognitive behavioral therapy (CBT), tinnitus retraining therapy (TRT), and pitch matching (PM)– treatment techniques are more effective when one's specific PTS is more accurately characterized (Davis, 2007; Landgrebe, 2012; Okamoto, 2009; Schaette, 2010; Wang, 2020). Tinnitus heterogeneity makes it difficult to generalize the experience of tinnitus and decreases treatment efficacy. There is a need for a clinical assay that can accurately characterize a patient's PTS to facilitate more effective sound therapy treatments.

This project aims to develop further a pre-clinical tinnitus characterization assay that can accurately and precisely estimate a patient's tinnitus, specifically those with a constant, non-tonal PTS. This assay– currently in development– is based on Reverse Correlation (RC), an established methodology to understand subjective cognitive representations better. RC has been used to study many aspects of visual perception and has also been incorporated into audio perceptions (Gosselin & Schyns, 2003; Mangini & Biederman, 2004; Dotsch & Todorov, 2012; Smith et al, 2012; Brinkman et al, 2017; Brimijoin et al, 2013). RC involves presenting a noise stimulus to a subject and testing which random stimuli evoke specific responses. At the outset of the present study, it was determined that, within the design space of RC-based tinnitus assays, the method of generating random stimuli for RC had the greatest need for further development.

Methodology

The overall steps of this RC tinnitus characterization assay are as follows:

- The generation method generates a stimulus.
- The test subject hears the stimulus.
- The subject response is collected.
- A number of these responses are then compiled and analyzed.

The experimental MATLAB protocol is evaluated on "healthy" (non-tinnitus) control patients in this pre-clinical setting to enable objective validation. Sample sounds from the American Tinnitus Association (ATA) were used as the "ground truth" target signal, and a stimulus sound generated using some technique. Subjects answered "Yes" if the stimulus sounded similar to the target sound and "No" if it did not.

We focused on optimizing the stimulus generation method of the tinnitus characterization method. In general, stimuli are created when a frequency spectrum of 100-13000 Hz is divided into a predetermined number of bins using the Mel scale (logarithmically scaled) (Umesh et al., 1999). The program systematically goes through all the bins and determines if and how they will be "filled," which translates to how that specific stimulus will treat the frequencies in that range.

Throughout this experiment, 4 stimulus generation methods were evaluated; we collected 500 responses from 4 test subjects for both the "roaring" and "buzzing" ATA sample sounds, recorded the time they took to complete the trials, and asked them a series of qualitative exit survey questions. Results were analyzed to determine the following stimulus generation method, and all methods were compared at the end. The stimulus generation methods were as follows:

A. **Uniform Prior:** 100 bins, 30 are randomly chosen to be filled at a flat, fixed power level
B. **Uniform Prior 8 Bin:** 8 bins, 3-7 are randomly chosen to be filled at a flat, fixed power
C. **Brimijoin:** 8 bins, 3-7 are systematically filled, with one of 6 possible flat powers
D. **Brimijoin Gaussian Smoothed:** 8 bins, 3-7 are systematically filled with one of 6 possible Gaussian-shaped powers.

Results

For Stimulus Generation A, the average time it took subjects to complete 500 trials was around 33.5 ± 12.5 minutes. This method had an average linear regression R-value of 0.126 ± 0.183 (R^2 value 0.045 ± 0.043). Subjects also reported low levels of confidence and understanding in exit survey questions.

For Stimulus Generation B, the average time it took subjects to complete 500 trials was around 30.0 ± 6.0 minutes. This method had an average linear regression R-value of 0.501 ± 0.358 (R^2 value 0.363 ± 0.346). Subjects reported slightly higher confidence and understanding-based exit survey questions.

For Stimulus Generation C, the average time it took subjects to complete 500 trials was around 24.8 ± 6.1 minutes. This method had an average linear regression R-value of 0.493 ± 0.346 (R^2 value 0.348 ± 0.310). Subjects also reported average to above-average in confidence and understanding-based exit survey questions.

For Stimulus Generation D, the average time it took subjects to complete 500 trials was around 28.6 ± 4.6 minutes. This method had an average linear regression R-value of 0.615 ± 0.240 (R^2 value 0.427 ± 0.282). Subjects reported the highest relative confidence and understanding via exit survey questions.

We performed an ANOVA and the Kruskal-Wallis significance test on several groupings of this data. Results showed the P value < 0.05, meaning that the null hypothesis could be rejected with 95% certainty for both the ANOVA and the Kruskal Wallis when all trials for each Stimulus Generation A-D were tested together. Post-hoc tests revealed that only Stimulus Generation A was statistically significantly different from Stimulus Generation D in both cases.

Discussion

Through this experiment, we determined that of the four stimulus generation methods evaluated, Stimulus Generation D: Brimijoin Gaussian Smoothed was the best method for this assay. This stimulus-generating method had the largest average linear regression R-values and the most positive exit survey results. It was also found to have a statistically significant difference from Stimulus Generation A via an ANOVA and a Kruskal Wallis and post hoc test.

Other considerations of this experiment are related to many test subjects completing trials for multiple stimulus generation methods. This means they were more familiar with the testing paradigm, which could have biased their qualitative exit survey results via carryover effects. Other sources of error include a lack of strict control of specific testing conditions; the environment in which test subjects took this assay was highly variable, meaning there was a substantial potential for environmental distractions. Subjective or age-related hearing differences amongst test subjects could have also affected subject signal reconstructions.

Authorship Page

Section(s)	Author	Editor
Executive Summary	Myah	Jacob
1. Introduction	Myah	Jacob
2.1 The Nature of Sound	Jacob	Myah
2.2 Tinnitus	Myah	Jacob
2.3 Reverse Correlation	Myah	Jacob
2.3.1-2 Reverse Correlation	Jacob	Myah
3.1 Project Design Approach	Jacob	Myah
3.2 State of Art Tinnitus Assay	Myah	Jacob
3.3 Data Collection	Myah	Jacob
4. Results	Myah	Jacob
4.1 SG A: Uniform Prior	Myah & Jacob	Myah & Jacob
4.2 SG B: Uniform Prior 8 Bin	Myah & Jacob	Myah & Jacob
4.3 SG C: Brimijoin	Myah & Jacob	Myah & Jacob
4.4 SG D: Brimijoin Gaussian Smoothed	Myah & Jacob	Myah & Jacob
4.5 Statistical Significance and Correlation Analysis	Myah	Jacob
5. Discussion	Myah	Jacob
6. Recommendations	Jacob	Myah
7. Conclusion	Jacob	Myah

Acknowledgements

We would like to thank our advisor Professor Adam Lammert, Ph.D. for helping us throughout the duration of this MQP project. His technical knowledge, academic professionalism, and unwavering support made the completion of this project possible. We would also like to thank Alec Hoyland and Nelson Barnett for the creation, development, and maintenance of the MATLAB pre-clinical tinnitus assay that facilitated our experimental methodology. This MQP also would not have been possible without the gracious participation of many anonymous test subjects who served as healthy controls when evaluating the preclinical tinnitus assay. Finally we would like to recognize Divya A. Chari, MD from UMass Memorial Health as a collaborator and clinical otolaryngology expert.

Supplemental Materials

List of Tables

List of Figures

List of Equations

Equation 1: Reconstruction of the Frequency Spectra

$$b\_hat = X^T y$$

where

b_hat = resulting reconstruction
X = stimuli frequency spectra
y = subject responses (1, -1)

Equation 2: Linear Reconstruction of the Frequency Spectra

$$b\_hat = (X^T X)^{-1} X^T y$$

where

b_hat = resulting reconstruction
X = stimuli frequency spectra
y = subject responses (1, -1)

Glossary of Terms

Word	Form	Definition
comorbidity	n.	a disease or condition that is often simultaneously experienced in relation to another condition
frequency spectra	n.	a range of frequencies (x) with associated power levels (y) composing a single stimulus sound, often graphically represented
habituation	n.	a process with the goal of achieving conditioned acceptance and disinterest
heterogeneity	n.	diverse manifestations of tinnitus, including (1) the sound and situations in which patients experience tinnitus, (2) causes and comorbidities, (3) reactions and distress, and (4) response to treatment (Cederroth, 2019)
idiopathic	adj.	when there is no apparent or identifiable cause (of a condition)
objective tinnitus	n.	tinnitus that has a tangible external cause and can often be heard by other people (ex: high blood pressure); as opposed to subjective tinnitus (Kaylie, 2022)
power	n.	the amount of energy in a sound; more specifically the decibel level assigned to a frequency(ies) in a frequency spectra
primary tinnitus	n.	idiopathic tinnitus; tinnitus that has seemingly no reason for onset; as opposed to secondary tinnitus (Tunkel, 2014)
psychoacoustic tinnitus spectrum (PTS)	n.	sounds, pitches, and volumes that patients perceive as tinnitus
secondary tinnitus	n.	tinnitus that has a specific identifiable cause; as opposed to idiopathic/ primary tinnitus (Tunkel, 2014)
stimulus	n.	a frequency spectra created using some stimulus generation method
subjective tinnitus	n.	tinnitus that has no identifiable external cause; as opposed to objective tinnitus (Kaylie, 2022)
tinnitus	n.	a persistent perception of sound without a corresponding external stimulus

1. Introduction

Tinnitus is a prevalent hearing condition affecting millions of Americans and people around the world. It can be characterized as a ringing, roaring, buzzing, or any other noise in the ear that is present without stimuli. It often ranges in severity but can lead to secondary symptoms such as insomnia, frustration, depression, and anxiety. Treatments often include reproducing the patient's sound and exposing them to it as a form of sound therapy. Tinnitus heterogeneity makes it extremely difficult to generalize the experience of those who present with this condition, as they can differ in how the patient perceives the sound. Critically, while methods for characterizing tinnitus sounds work well for patients who experience tonal (e.g., "ringing") tinnitus, current methods are inadequate for characterizing non-tonal tinnitus sounds. This gap makes it difficult to precisely replicate the unique sound one's tinnitus creates for many patients, decreasing the efficacy of any subsequent sound therapy treatment. This highlights the need to develop methods to characterize non-tonal tinnitus accurately.

Recent developments towards accurately characterizing one's non-tonal psychoacoustic tinnitus spectrum (PTS) have been based on Reverse Correlation (RC). Reverse Correlation (RC) is an established method for characterizing one's internal representation of some sensory perception; it has historically been used in visual applications– such as estimating the representation of the letter "S" or an "angry" facial expression– but has since been incorporated into auditory perceptions. RC involves subjecting a participant to randomized stimuli and asking them to indicate "Yes" if it is similar to their perception or "No" if it is not similar to their perception. RC is a candidate for characterizing non-tonal tinnitus because it can recreate a patient's PTS with greater accuracy and fewer biases than other methods currently in practice, such as Pitch Matching (PM) methods. This increased accuracy has the advantage of being used for more effective sound therapy treatments, specifically customized to a patient's PTS. Early feasibility studies have indicated that RC has potential for tinnitus applications; however, it has not been optimized to produce the best results. The method for generating these "random" stimuli could be more efficient to increase the characterization's effectiveness, accuracy, and precision.

This project aims to develop further an assay that can accurately and precisely estimate a patient's tinnitus, specifically those with a constant, non-tonal psychoacoustic tinnitus spectrum (PTS), thereby enabling more effective methods of treating tinnitus. This will be achieved by examining, evaluating, and optimizing various stimulus-generation approaches and techniques.

The long-term vision for this research is that it may be helpful in a clinical setting by developing an accurate synthesization of a tinnitus patient's specific PTS for more effective sound therapy. In the future, a tinnitus patient may undergo a computer-based tinnitus characterization assay with controlled patient instructions, stimulus generation, data processing, and a general experimental paradigm. This will output a PTS reconstruction suitable for sound therapy and improve the treatment options for patients with non-tonal tinnitus. It will also streamline clinical sound therapy protocol overall.

2. Literature Review

As humans, we interact with the world in various ways; snoozing an alarm, sniffing if the milk has gone sour, eating breakfast cereal, reading the morning paper, or shaking hands with a coworker. We are constantly taking in input from our surroundings through the use of the five senses: hearing, smell, taste, sight, and touch. These senses are invaluable, and as such, they are missed when absent or impaired. Hearing is of particular interest to us in this present work as it relates to the condition of tinnitus— a persistent internally-generated auditory phenomenon.

In this chapter, we will begin by discussing the nature of sound and how it is made, characterized, and perceived. Next, we will move on to exploring tinnitus in greater detail. This entails highlighting the different types of tinnitus and the scope of which is considered. Other relevant information, such as prevalence, demographics, comorbidities, and causes, will be outlined before we move on to the current research and treatments. The advantages and challenges of these treatments will be described as well. Finally, we will turn our attention to the analytical side of this research by discussing the idea of reverse correlation. We will describe what it is, its types, and how it is typically used. From there, we can entertain the future potential for increased efficiency of this analysis method.

2.1 The Nature of Sound

Sound, as it is commonly understood, is a result of the vibrations of objects. The vibrations of these objects are transferred through a series of pressure changes in the surrounding medium; in the context of human hearing, the relevant medium is most frequently air. The interpretation of these pressure changes in the ear is what most people interpret as sound.

2.1.1 Components of a Waveform

To accurately describe sound visually and mathematically, researchers developed a general equation for describing the pressure changes associated with sinusoidal vibrations, representing the simplest sound. The equation exists in the form $\Sigma\, \alpha\, cos(ft + \phi)$ where α is the volume, f is the pitch, and ϕ is the phase (Moore, 2013). In a graphical sense, α is the amplitude, or height, of the signal, f is the frequency, which is the inverse of the wavelength, and ϕ is how much the graph is shifted horizontally. These three variables can be adjusted to create variations of tones, which look like Figure 1 when graphed.

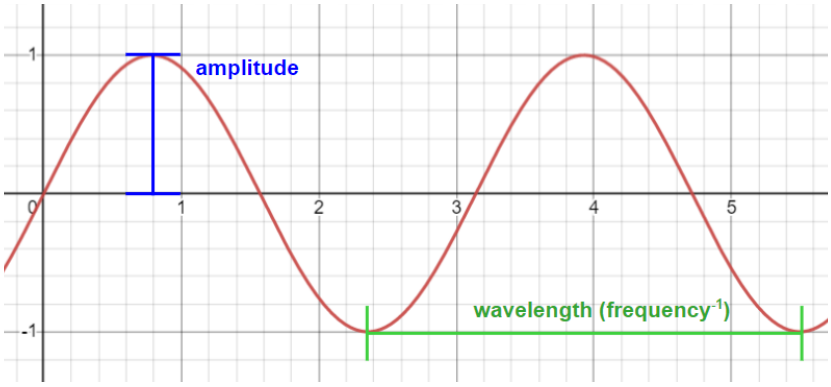


Figure 1: Sound wave visualization.

These tones are simple, and are commonly referred to as pure tones. Even though the graph varies from negative to positive values, the ear will interpret this as a flat noise, such as a beep. However, most sounds are not pure tones and must be represented differently. One example would be speech. Speech can be characterized by a series of fluctuations in the frequency and amplitude. However, such examples are highly complicated to graph and interpret. Simpler sounds, however, are helpful to this study. Such sounds include white noise or other replications of tinnitus tones. These are achieved by adding the product of several pure tones. These can create sinusoid graphs that no longer look simple but have great degrees of variation. This variation is what allows speech to have different syllables and sounds.

2.1.2 Anatomy of the Human Ear

It is helpful to have a general understanding of auditory anatomy and how each part of the ear interacts with others to appreciate the complexity of tinnitus. An image detailing the anatomy of the ear can be seen in Figure 2 below.

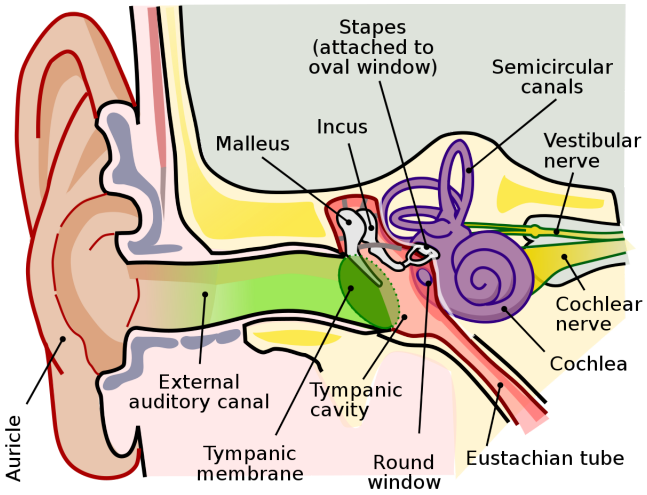


Figure 2: Outer, middle, and inner ear anatomy (Brockmann, 2005; licensed under CC BY 2.5).

To begin to understand, it is best to start with the outer ear. The outer ear includes the auricle, the external auditory canal, and the outer layer of the tympanic membrane. The auricle is the cartilage portion of the ear located outside the head, and the tympanic membrane is the outer layer of the eardrum. The primary function of the outer ear is to funnel sound waves into the middle and inner ear.

The middle ear comprises the eardrum, cavity, and three small bones. The eardrum converts sound energy into mechanical energy to continue in the ear. The bones are called the malleus, the incus, and the stapes, and their purpose is to propagate the sound further into the ear.

The inner ear is the most crucial part of the ear; it converts the mechanical vibrations from the middle ear into the nerve signals most people hear as sounds. Within the inner ear are the cochlea and the auditory nerve. The cochlea is a spiral structure that coils within the inner ear and is covered in microscopic organs called hair cell receptors. These hair cell receptors translate the mechanical energy in the ear into electrochemical signals that the auditory nerve makes into sound for most people (Anatomy of the Ear, n.d.; *Parts and Components of Human Ear and Their Functions*, n.d.; Perry, 2021).

2.2 Tinnitus

Tinnitus is documented as a possible symptom associated with over three-quarters of all hearing ailments (Kaylie, 2022). It relates to how one perceives an auditory input in the absence of an external stimulus. There is a wide variety of sounds, pitches, and loudnesses that patients report this 'ghost noise' as sounding like, the whole of which is referred to as the *psychoacoustic tinnitus spectrum* (PTS). There is also a diverse range of situations in which these sounds make themselves known to patients: sound-induced or silence-induced. These diverse manifestations of tinnitus are described as tinnitus *heterogeneity* (Cederroth, 2019). Tinnitus is very common, with 3%-30% of people affected worldwide (Bhatt, 2016). Causes of tinnitus range from physical, such as repeated exposure to loud noises, to neurological abnormalities. Treatment options, like sound exposure therapy, can be administered through stimulation devices or made more accessible through smartphone apps (Tunkel, 2014). Future research aims to improve the efficacy, accessibility, and efficiency of the currently available diagnostic and treatment options to maximize the benefits to patients experiencing diverse PTSs.

2.2.1 Understanding Tinnitus

Tinnitus is when a patient perceives some form of auditory sensation without an attributed stimulus. In other words, it is when someone hears a sound without an external source. Tinnitus is not a disease but can be a symptom of other diseases. It is as diverse as it is prevalent, meaning that tinnitus heterogeneity is such that the condition affects each person differently (NIDCD, 2017). Tinnitus heterogeneity results in differences in everything from how the sound is perceived by the patient, to causes, to related implications and comorbidities, to the level of

impact on one's daily life, to how much improvement results from treatment (Cederroth, 2019). A representation of the facets of tinnitus heterogeneity can be seen in Figure 3 below.

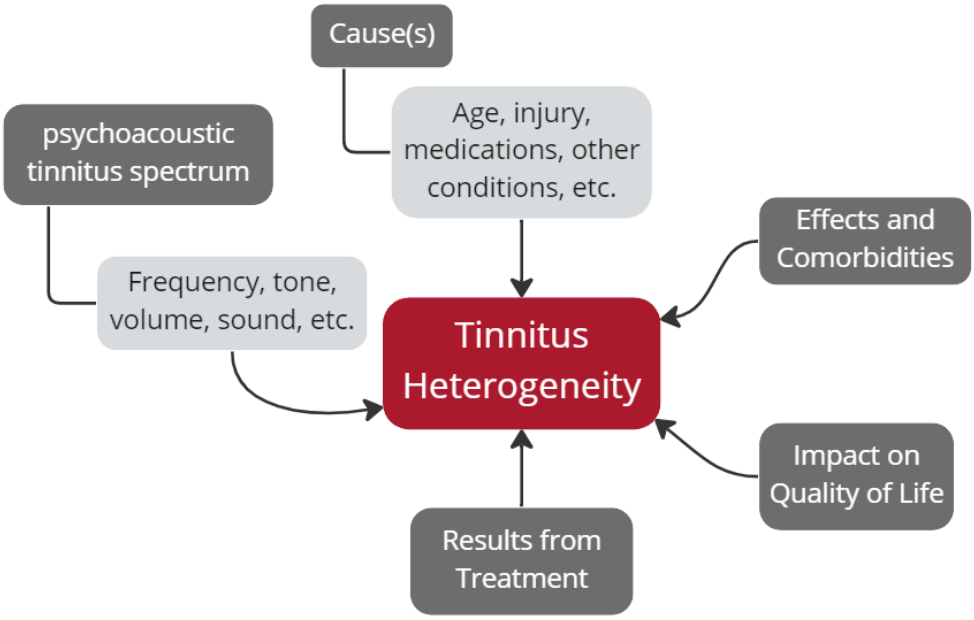


Figure 3: Understanding tinnitus heterogeneity (Cederroth, 2019; Kaylie, 2022; NIDCD, 2017).

Patients can experience tinnitus so mild that it is barely noticeable to so loud that it drowns out external sounds and inhibits one's day-to-day life. This noise not only ranges in volume but also pitch. Tinnitus can be described as ringing, roaring, buzzing, humming, hissing, chirping, clicking, whistling, whirling, and whooshing, among other things (Kaylie, 2022). Additionally, the times one experiences tinnitus onset is different for each patient as well; some experience tinnitus in near-silent situations, like when they are trying to sleep, and some experience noise-induced tinnitus, like when in a crowded public area (Shekhawat, 2014). Some people also experience constant tinnitus, as opposed to intermittent or situational.

2.2.2 Causes and Prevalence of Tinnitus

Tinnitus of varying degrees and severities is extremely common in many populations. In the United States of America, it is estimated that 8%-25.3% of people experience some form of tinnitus. This amounts to roughly 50 million people in the United States of America (Kaylie, 2022; Tunkel, 2014). Similarly, about 3%-30% of people worldwide are believed to experience tinnitus (Sanchez, 2004). A study published by the *JAMA Otolaryngology* journal found that of the people who suffer from tinnitus, “27% had symptoms for longer than 15 years, and 36% had nearly constant symptoms” (Bhatt, 2016). The study continued to conclude that “7.2% [of sufferers] reported their tinnitus as a big or a very big problem compared with 41.6% who reported it as a small problem,” (Bhatt, 2016).

Tinnitus can have several causes ranging from prolonged exposure to high decibel levels to physical trauma to the inner ear. Tinnitus can also indicate other physiological events, such as

high blood pressure or muscle contractions in areas surrounding the ear. Some patients experience pulsatile tinnitus, in which the noise is rhythmic, usually related to one's heartbeat (NIDCD, 2017). Tinnitus with an identifiable cause is known as *secondary tinnitus*, whereas *idiopathic* tinnitus– or seemingly that without cause– is known as *primary tinnitus* (Tunkel, 2014). Because of the variety of situations that can lead to tinnitus, a doctor will likely examine a patient to determine if there are any ongoing physiological causes. Some of these identifiable causes include natural hearing loss attributed to age, trauma to the inner ear leading to the bending or breaking of the inner hair cells, head, neck, or and brain injuries leading to neurological issues, high blood pressure resulting from nicotine, alcohol, caffeine, or strenuous physical activity, and ototoxic medications that cause concentration-dependent tinnitus (Yew, 2014). The tests used to identify these causes range from audiological exams to CT scans and MRI imaging tests to other lab tests like blood work (Biswas, 2021; Tunkel, 2014).

Due to the specific cause of tinnitus, certain populations are more at risk for experiencing this condition. In some cases, tinnitus affects people as a natural consequence of aging, meaning that adults are the most likely sufferers of this condition (Nondahl, 2007). People whose occupation involves being around loud noises are more likely to experience tinnitus, such as construction workers, soldiers, singers, audio technicians, and pilots.

Tinnitus is associated with other complications and correlated ailments. Symptoms directly related to tinnitus include trouble sleeping, attention deficit, trouble focusing, and migraines. Other more chronic symptoms commonly present in conjunction with tinnitus include stress, memory impairment, and irritability or anxiety (Bhatt, 2016).

2.2.3 Treatment and Research of Tinnitus

Tinnitus heterogeneity, or the diverse spectrum across which tinnitus manifests and behaves, means various treatment options exist for this condition. Firstly, a healthcare professional will attempt to identify the underlying cause of the patient's tinnitus. This can be done in various ways, such as an audiological exam, a CT scan or an MRI, blood tests, or even a close inspection of the patient's drug regime (Baguley, 2013; Tunkel, 2014; Yew, 2014). Once the cause of the patient's tinnitus is ascertained, an appropriate treatment method can be selected.

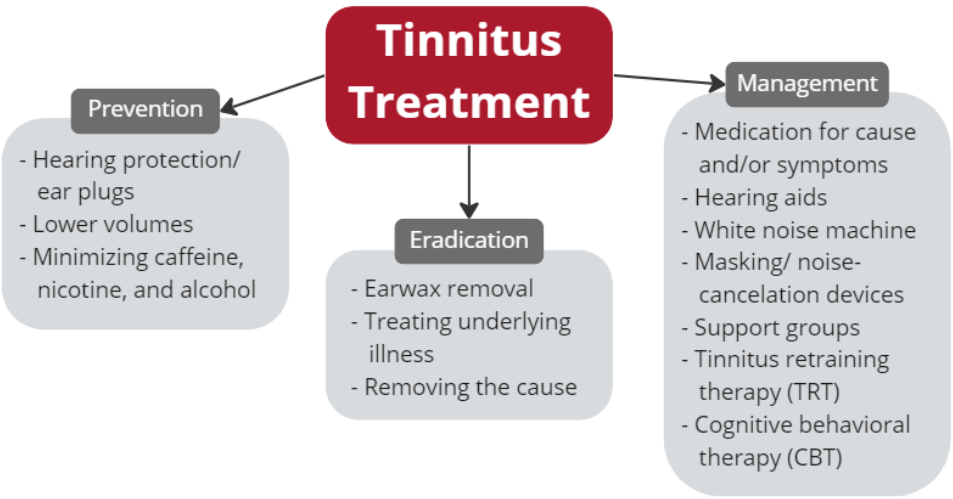


Figure 4: Tinnitus treatment approaches (Baguley, 2013; Kaylie, 2022; Larem, 2021; NIDCD, 2017; Tunkel, 2014; Yew, 2014).

We can define tinnitus treatment methods, consistent across various sources, into three broad categories: prevention, eradication, and management, as summarized in Figure 4. Prevention is the proactive inhibition of tinnitus onset and progression. Methods of preventative tinnitus treatment include avoiding listening to music, TV, radio, and other audio sources at a loud volume and wearing hearing protection– such as ear plugs– when loud environments cannot be avoided (Kaylie, 2022; Larem, 2021; NIDCD, 2017; Tunkel, 2014). A benefit of this type of treatment is that it can reduce the later severity of tinnitus and age-related hearing loss. On the other hand, prevention is not as useful for people already experiencing moderate to severe tinnitus.

Eradication treatments are those we view as attempting to curtail one's tinnitus experience by eliminating the perceived direct cause. These eradicative treatments are typically used where physical blockage or infection is the cause of one's tinnitus or where tinnitus is a known side effect of a disease or medication. Methods of treatment include removal of earwax or flushing of the ear canal, changing the patient's medications, or treating the underlying condition, such as taking blood pressure stabilizers if experiencing hypertension or muscle relaxers if experiencing localized spasms (Baguley, 2013; Kaylie, 2022; Larem, 2021; NIDCD, 2017; Tunkel, 2014; Yew, 2014). A benefit of eradication treatments is that they often provide instant, if not fast, relief for tinnitus sufferers. A clinician can also relatively quickly achieve this in a few visits. Nevertheless, these methods only lend themselves to specific causes of tinnitus and therefore are not helpful for age-related or idiopathic tinnitus.

The final type of tinnitus treatment options can be considered for tinnitus management. Management is often the best option for the idiopathic onset of tinnitus that cannot be prevented or eradicated. The easiest management technique is for a doctor to prescribe medications to treat the effects and complications of tinnitus, such as anxiety, insomnia, or attention deficit. This can help minimize the burden of the condition. Another management technique is using an external

device, such as hearing aids, masking devices, or white noise machines. Hearing aids can help amplify external noises to be louder than the patient's tinnitus; masking devices can be worn to attempt to cancel out the sound of the patient's tinnitus and make it less perceivable; and white noise machines can help displace the patient's attention from their tinnitus to an external sound, especially in the case of silence-induced onset (Baguley, 2013; Kaylie, 2022; Larem, 2021; NIDCD, 2017; Tunkel, 2014; Yew, 2014). The benefit of these tinnitus treatments is that they can help ease the burden of tinnitus that is otherwise unresponsive to other treatments. Nevertheless, the downside to this type of treatment is that it can be costly, may require the advice or supervision of a medical professional, and may take some time to experience lasting or noticeable effects.

Other more rigorous forms of tinnitus management treatments are those involving sound therapy. Sound therapy can come in different forms, ranging from cognitive behavioral therapy (CBT) to tinnitus retraining therapy (TRT), to Heidelberg neuro-music therapy (HNMT), to Tailor-made notched music training (TMNMT), to Tinnitus pitch-matched therapy (PM), and several others (Wang, 2020). These techniques generally involve habituating a patient's tinnitus responses, or the conditioned acceptance and disinterest. They aim to convert those who experience severe and bothersome tinnitus to those who experience tinnitus and whose lives are not actively impaired (Jastreboff, 2011). The most common and most accessible types of sound therapy are CBT, TRT, and PM. CBT is a treatment technique that reduces the patient's negative response to tinnitus. Supervised by a clinician, this treatment challenges one's automatic response and aims to retrain these negative automatic responses to be more neutral (Zenner, 2013). TRT combines a comprehensive medical history analysis, sound therapy, and professional counseling. Similarly, this treatment aims to decrease the patient's perception of tinnitus and improve their quality of life (Grewal, 2014).

Another common sound therapy component is pitch matching, which can be used by itself or as a component of CBT and TRT. This therapeutic technique involves determining a sound pitch that best resembles the patient's tinnitus experience, and as such, this method is most effective in tonal tinnitus cases. PM aims to alter the patient's conditioned response to their tinnitus perception and minimize its adverse effects. Recently, different PM methods have been tested, hoping to increase the efficiency and accuracy of characterizing one's internal auditory perception (Henry, 2004). This has been done with various techniques ranging from manual to computer-generated testing procedures.

The benefits of management tinnitus treatments are that they can provide some relief for tinnitus sufferers who would otherwise not be able to receive the other forms of treatment. On the other hand, these treatment options are relatively time-consuming, require multiple sessions with an experienced clinician, and are not guaranteed results. Management treatment options for tinnitus have been recorded as anywhere from 0-60% effective (Jastreboff, 2011).

While there are several available sound therapy techniques– including CBT, TRT, and PM– each is more effective when they more closely resemble one's specific psychoacoustic tinnitus spectrum (PTS). Studies have shown decreased tinnitus loudness and decreased

tinnitus-related discomfort in patients treated with more accurate tinnitus characterizations (Davis, 2007; Landgrebe, 2012; Okamoto, 2009; Schaette, 2010; Wang, 2020). Historically, the need for such specificity of tinnitus sound therapies– combined with the vast heterogeneity– made the condition extremely difficult to treat. Better tinnitus characterization and replication methods have increased the efficacy of CBT, TRT, and PM treatments.

These treatment options are utilized and proven effective, but there is still ongoing research, and the specific delivery of these methods varies. Current efforts are geared towards improving the lives of people suffering from tinnitus through the most straightforward and accessible avenues. As a result, different treatment options– specifically management treatments like CBT– have recently been integrated into smartphone apps (Mehdi, 2020). The increased efficacy of tinnitus characterization and treatment methods is augmented by the increased accessibility of tinnitus treatment for those afflicted.

2.3 Reverse Correlation

As discussed, the efficacy of tinnitus treatments drastically improves with better characterization techniques. This is seen through patients reporting decreased loudness in their tinnitus and less distress due to their condition (Schaette, 2010). Current characterization methods are dominated by pitch matching (PM) techniques. The PM process usually involves playing some sound for a patient and asking them to give feedback on how accurately it mimics their tinnitus. PM has three components: defining the threshold, identifying the loudness, and matching the pitch (Henry, 2004). Given the inputs of this technique, pitch matching is generally only effective for patients who experience constant tonal tinnitus.

Reverse correlation is a methodology developed to understand better the internal architecture related to a subject within someone's mind. This technique was developed primarily for psychological applications in which the perception of a research subject or population is questioned. It involves applying noise to a neutral stimulus and testing which random stimuli evoke specific responses. The Gosselin and Schyns (2003) study was the first significant validation of this technique. This study showed participants a visual stimulus of randomly generated white static noise. They were instructed that half of the images contained a black letter “S” on a white background– underneath the applied noise– and asked to indicate which images they believed contained the “S.” After tens of thousands of trials, the researchers were able to compile participant responses and yield a visualization of what the patient perceived as the described “S.” This experiment has extremely practical implications in reconstructing unobservable internal sensory perceptions. Since the publishing of this paper, RC has been widely used to study many aspects of visual perception and has been incorporated to apply to audio perceptions as well.

2.3.1 Historical Applications

Although reverse correlation has been applied to several experimental paradigms, it remains under constant development. It has been proven effective in experiments generally regarding perception, whether it be visual or auditory. After the Gosselin and Schyns (2003) "S" example mentioned above, reverse correlation has also been used in several other experiments. The subsequent wave of experiments based on RC continued to explore the methodology applied to visual perceptions. A study by Mangini and Biederman (2004) used RC to reconstruct the visual participants' internal representations of more abstract concepts than simply the letter "S"; they instructed participants to indicate if noisy 50:50 composite images of faces were "male or female," "happy or unhappy," and "Tom Cruise or John Travolta." From the responses of these participants, researchers were able to reconstruct these six classifications, as shown in Figure 5 below.

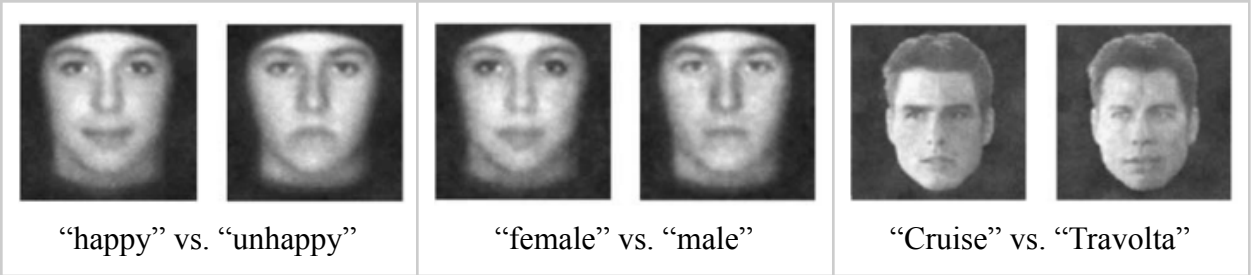


Figure 5: Reconstructions of internal visual representations using RC (Mangini and Biederman, 2004).

Subsequent studies, such as one by Dotsch and Todorov (2012) and another by Smith et al. (2012), continued building on evidence that RC could apply to abstract and subjective ideas. These studies focused on applying RC to data concerning the presence or absence of a face and the presence or absence of specific characteristics such as "trustworthy" or "untrustworthy" and "dominant" or "submissive." Based on previous work in the field, Brinkman et al. (2017) have postulated that RC could be used to determine what factors are most significant in perceptions of certain qualities.

The next substantial development in the vein of reverse correlation studies was when Brimijoin et al (2013) applied this methodology to audio perceptions. The subjects were asked to determine if they perceived a sound more like the vowel "a" or "i" and respond accordingly. Through this, the researchers created spectrograms for each group of responses. The spectrograms that were created matched the classic vowel formants of each letter. This study found that the subjects' perceptions of each letter were not significantly affected by the noise present in the stimuli, and researchers were confident that RC might be applied to other cases of audiological perceptions.

Just as Mangini and Biederman (2004) and Dotsch and Todorov (2012) expanded on the visual RC work of Gosselin and Schyns (2003), a study by Ponsot et al. (2018) built upon the

audiological RC developments of Brimijoin et al. (2013) and Brinkman et al. (2017). Ponsot et al. (2018) sought to find how pitch affects perception, specifically in the capacities of trustworthiness and dominance. The experiment found merit in using reverse correlation in the scope of audiology in more abstract or subjective applications.

2.3.2 Compressive Sensing and Efficiency

One of the greatest barriers to reverse correlation is the massive amount of data that has been required to conduct these studies to a significant degree; the "S" experiment described required 20,000 responses from each test subject. This is due to the inefficient methods used in the past to analyze the data given (Gosselin and Schyns, 2003). However, there have been efforts to reduce the necessity for collecting so much data. There have been attempts in the past to achieve improved efficiency, but, unfortunately, at the cost of the accuracy of the experiment.

Previous attempts to reduce the amount of stimuli-response pairs needed have biased the result. In one study, the initial sample was slightly biased to increase the likelihood of the subjects responding positively (Moon et al, 2020). Another solution to increase the efficiency of reverse correlation has been to impose constraints on the data, and one example is putting it through a low-pass filter. This decreases the amount of high-frequency data present but relies on high-frequency data to be unimportant in the analysis.

3. Methodology

The research in this project was done within the context of a larger project, creating a Tinnitus characterization assay. Given that this portion of the research aimed to validate the stimulus generation methods, the experimental method used consisted of an AX paradigm that compared data from multiple iterations that used different stimulus generation methods. Most of the data collected was quantitative and allowed for direct comparison between stimulus generation methods. This data collection was researcher guided and overseen. The researcher would give direct instructions, answer questions, and provide information to the test subject as necessary or requested. This methodology allowed for the most complete and accurate gathering of information possible.

3.1 Project Design Approach

The research was structured using a design space described in section 3.1.1. The design space is a concept used frequently in research like this, although usually without it being realized or acknowledged. The details of each dimension of the design space are what make each experiment unique. Section 3.1.2 describes how the specific focus of the research was decided and what was deemed most important to this study. The limited time and test subjects available forced the research to prioritize specific testing methods; this section elaborates on the reasoning for the choices made.

3.1.1 The Design Space

The proposed assay's design space comprises four aspects: patient instructions, stimulus generation method, reconstruction method, and experimental paradigm. The literature on reverse correlation reveals that these elements vary with the intended application area. For example, the stimulus generation method varies per the chosen perceptual domain (e.g., vision, hearing), target representation (e.g., tones, phonemes), and preferred level of detail and accuracy in the reconstruction. Therefore, effectively designing an assay for tinnitus based on reverse correlation will depend on carefully selecting the assay's parameters along each of these four design dimensions.

The first dimension of the design space is patient instruction. The goal of patient instruction is to instruct the test subject on how to take the test and do so consistently. Ensuring that patient instruction is consistent is critical in producing valid test results. A typical patient instruction might sound as follows:

"For this project, you will listen to two sounds and respond according to how similar you think they sound. The first time you hear it will be the control sound; this will remain constant throughout the test. The second sound will be the generated sound. If you think the two sounds are similar, give an affirmative answer. If you think they sound different, give a negative answer. The project's graphical user interface (GUI) is on the screen in front of you. As you can see indicated on the screen, the "F" key indicates a positive answer. Similarly, the "J" key gives a negative answer. This test will be run in five blocks of 100 trials with a break as needed between each block. Do you have any questions at this time?"

The next element of the design space was stimulus generation. This is the element that was tested primarily in the current project. The stimulus generation method for creating the test sound still needed to be optimized. Therefore it was essential to look into what method would be optimal. A variety of stimulus-generation methods were considered. However, given the feasibility of how many trials could be performed, only four stimulus-generation methods were tested. These stimulus-generation methods were based on various distribution functions. These varied from uniform distributions to Gaussian distributions. This will be discussed further in section 3.1.2.

The third element of the design space was the reconstruction. The reconstruction of the frequency spectra was done using the equation:

$$b\_hat = X^T y \qquad [1]$$

where b_hat is the reconstruction, X is the stimuli and, y is the responses. The test subject's responses, "yeses" and "nos," were compared, and the difference between them was found. The difference between the subject's positive and negative responses was used as their final product. Alternative reconstruction methods include:

$$b\_hat = (X^T X)^{-1} X^T y \qquad [2]$$

which is equivalent to linear regression. Eq 1 is a restricted form of Eq 2, assuming that stimuli are uncorrelated. Another reconstruction technique is compressive sensing (Roop, 2022). Reconstruction consistent with Eq 1 is the most common method in the literature and was therefore adopted here.

The fourth element of the design space was the experimental paradigm. The paradigm used in the experiment was an AX paradigm. An AX paradigm typically uses a control stimulus followed by a second experimental stimulus. This allowed for an experience comparable to an actual tinnitus patient, where they would compare a sound to their tinnitus tone. A prominent alternative to the AX paradigm sometimes used in reverse correlation experiments is a two-alternative forced choice (i.e., 2AFC) experiment, in which subjects compare two sounds and select the best one. The AX paradigm is more commonly represented in the literature than 2AFC in reverse correlation experiments and was therefore selected here.

Another aspect of the paradigm mentioned previously was that the design for this experiment was iterative. Each new stimulus generation method was tested independently and then compared to the previous methods to determine if it improved and if that avenue should continue to be explored.

3.1.2 Defining and Limiting the Scope

As mentioned earlier, the focus of this research was to find a stimulus generation method that was optimal for finding a tinnitus patient's true tinnitus tone and, as mentioned in section 3.1.1, there were a plethora of potential options for stimulus generation methods, such as using binary distributions, uniform distributions, Gaussian distributions, and others. The first factor that was considered was the timeline. The amount of time given for the project was approximated and a timeline was created with that information. This timeline showed that there would be time to test four stimulus-generation methods. The iterative design was decided at this point, and it began with the stimulus-generation method that had shown the most promising results. After this point, there was no longer any data to compare to the results of the first trials, so the stimulus-generation methods that would be tested next were decided in conversation with the project advisor.

The new stimulus generation methods were chosen based on a few factors which varied each time data were collected. The first factor that was addressed was the size of the bins. Initially, the spectrum was divided into 100 bins. After testing this method, it was noted that there was minimal variability from stimulus to stimulus because of such a high number of total and filled bins. The sounds became normalized and sounded similar to white noise. Once the amount of total and filled bins was reduced and it was easier for subjects to discern between stimuli, the power of each bin was considered. Having a variable power level in each bin could be more valuable than having binary power levels, so that was the next change implemented. For the final method, the unnatural nature of each sound was considered. The rectangular shape of each bin likely caused this. Therefore, if the shape of each bin could be smoothed, then the sound

would sound more natural to the test subjects. This was addressed by giving the bins a Gaussian shape.

3.2 State-of-the-Art Tinnitus Assay

The test program used for this experiment was written in MATLAB. It can be accessed via GitHub at the following address https://github.com/alec-hoyland/tinnitus-reconstruction. Future iterations of this program may be available through Julia.

From a conceptual standpoint, the overall steps of this tinnitus assay program are as follows:

- A stimulus is generated
- The stimulus is played for the test subject
- The subject response is collected
- A number of these responses are then compiled and analyzed.

A diagram of this process can be found in Figure 6 below. This core process remains unchanged regardless of what perturbations are made to the design space.

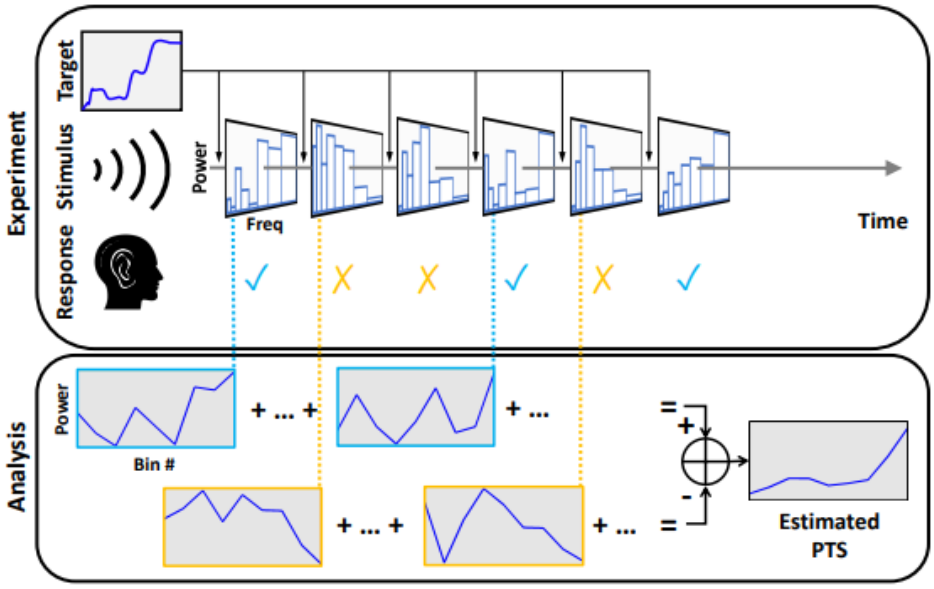


Figure 6: Experimental protocol overview (Hoyland et al, 2023).

3.2.1 Stimulus Creation

The general process of stimulus creation is as follows: a frequency spectrum of 100 - 13000 Hz is divided into a predetermined number of bins. This number of bins changes based on the stimulus generation method. The bins are divided evenly using the Mel scale (Umesh et al., 1999). The Mel scale converts frequencies into a scale that represents the perceived difference in pitch (i.e., the difference between 0 mel and 250 mel is comparable to the difference between 2750 mel and 3000 mel).

Next, the program systematically goes through all the bins and determines if and how they will be "filled," which translates to how that specific stimulus will treat the frequencies in that range. If a bin is "filled," all of the frequencies in its corresponding range have an increased relative decibel level compared to bins without power. An example frequency spectrum with 8 "bins" can be seen in Figure 7 below.

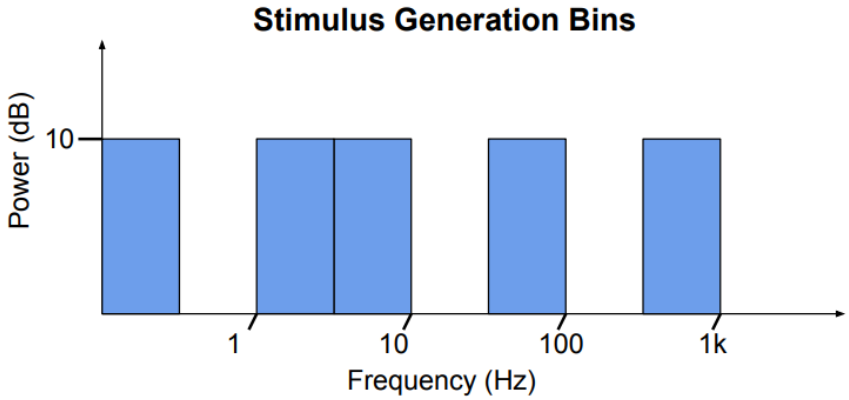


Figure 7: Example binned frequency spectrum (note that the x-axis is a log scale).

The stimulus generation methods differ in how they choose and fill bins. Some examples of stimulus generation methods and the way they handle these aspects can be found in Table 1 below. The total bin number, and maximum and minimum number of bins filled, can also be changed independently of these factors.

Table 1: Stimulus Generation Methods and Descriptions.

Name	Bin Choice	Bin Power	Bin Shape
Bernoulli Stimulus Generation	50:50 chance	Binary power	Flat
Brimijoin Stimulus Generation	Random distribution	6 possible powers	Flat
Gaussian Noise No Bins Stimulus Generation	Gaussian by each freq	Binary power	Flat
Gaussian Noise Stimulus Generation	Gaussian distribution	Binary power	Flat
Gaussian Prior Stimulus Generation	Gaussian distribution	Binary power	Flat
Uniform Prior Stimulus Generation	Randomly chosen	Binary power	Flat
Brimijoin Gaussian Smoothed Stimulus Generation	Random distribution	6 possible powers	Gaussian

This experiment program is such that every test subject needs a config file to set the parameters for the experiment. This permits features of the experiment to be modularly changed, without rewriting the code, just by selecting the proper config file. Namely, this allows one to: select where trial data would be saved, input the subject ID, change the stimulus generation method, choose a target signal, decide the number of bins, set a minimum and maximum number of bins, and dictate the stimulus generation method. An example config file can be found in Appendix 9.1. Config files also allow the number of trials per block and the number of blocks to be set, determining how often the test subject is prompted to take a break and when all trials have concluded.

3.2.2 Test Subject Interface

The American Tinnitus Association (ATA) has provided several audio file examples of non-tonal tinnitus, including ones described as buzzing, roaring, screeching, electric, static, and a tea kettle (Listen to Sample Tinnitus Sounds, 2022). This is useful in validating this experimental program because it allows the methodology to be tested with healthy control subjects. In other words, a healthy control could perform trials where they hear (1) the "ground truth" target signal, being one of the ATA sample sounds, and (2) a stimulus sound generated using some technique. They can then answer "Yes" if the stimulus sound is similar to the target sound and "No" if it is not (see Section 3.1.1 for an explanation of subject instructions).

This input from test subjects was acquired through the MATLAB program, which, when run, displays the screens below. Figure 8 shows the initial instructional screen displayed at the start of the test.

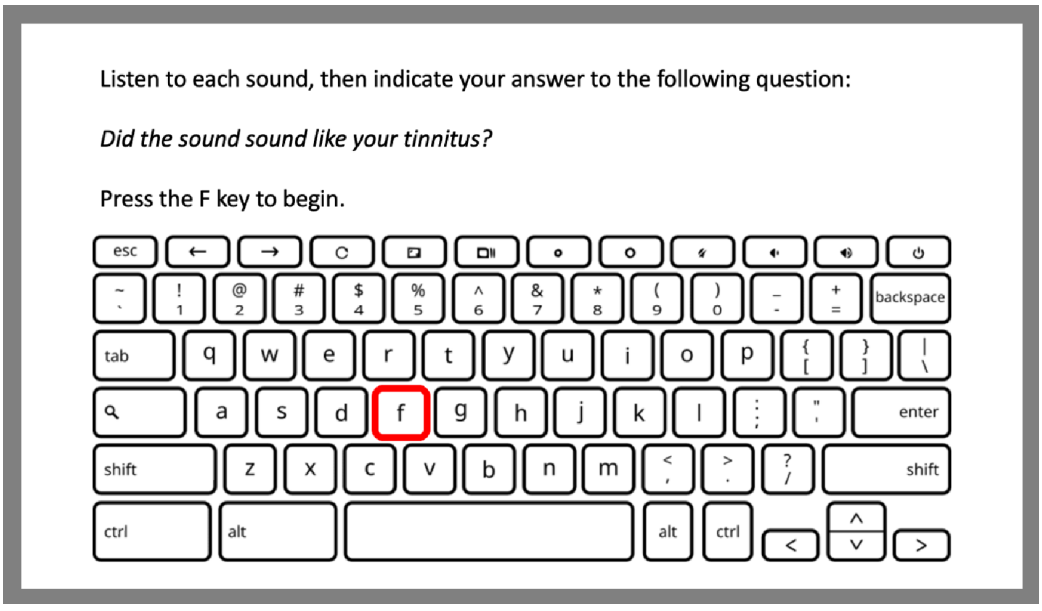


Figure 8: MATLAB program introductory GUI.

Once the subject presses “F” and trials begin, the screen shown in Figure 9 is presented. This screen reminds users how to input their answers in response to the different stimuli generated by the assay.

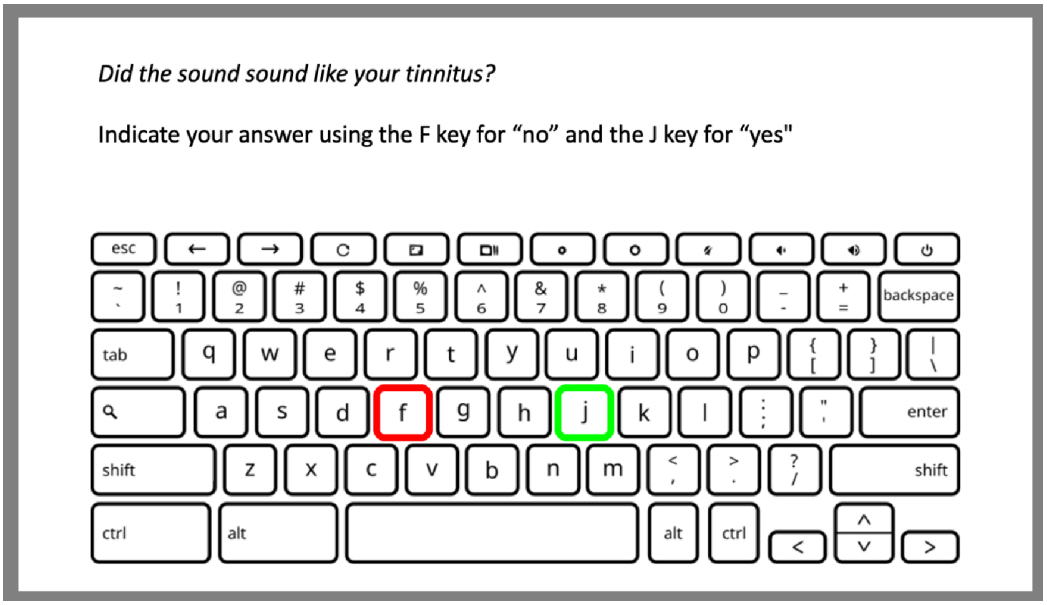


Figure 9: MATLAB program test trial GUI.

From the test subject’s perspective, this feedback process mimics the procedure that an actual tinnitus patient may undergo in a future clinical setting. Tinnitus patients would compare the presented stimuli to their own internal, perceptual representation of their tinnitus sound. From a research standpoint, this modified, healthy-control protocol has the added benefit of having an objective target signal, meaning that the performance and accuracy of different experimental components can be examined and compared.

3.2.3 Response Compilation and Analysis

After collecting several subject responses, they are compiled and analyzed to produce a synthesization of the subject’s perception of the target signal. Responses from different blocks are stored in different files that can be located and called by their hash identifiers when performing analysis. An analysis is done by effectively averaging the powers of bins along the frequency spectra from all trials. The reconstruction of the frequency spectra abides by the following equation:

$$\text{b\_hat} = X^T y \qquad [1]$$

where b_hat = resulting reconstruction
X = stimuli frequency spectra
y = subject responses (1, -1)

Reverse correlation allows the subject's "Yes" and "No" responses to be compared. All "Yes" responses and all "No" responses are summed. The "No" responses are then subtracted from the "Yes" responses, and the difference between these subject responses is used to create the reconstruction frequency spectra.

In addition to a frequency reconstruction generated via subject responses, statistical analysis of these responses is also performed. To do this, the target signal is binned based on the parameters of the stimulus generation method. In other words, if the stimulus generation is a Uniform Prior 8-bin method, the target signal will be approximated using eight bins filled with a flat, uniform power. The code used to approximate a given target signal using the Brimijoin Gaussian Smoothed Stimulus Generation method can be found in Appendix 9.4 (see Section 3.3.3 for a more detailed description of this stimulus generation method). The analysis process compares this "perfect" target signal approximation to the subject reconstruction frequency spectra. This yields an R correlation coefficient of the reconstruction compared to the "binned" target signal.

3.3 Data Collection

After the design space was outlined, with a narrowed focus on stimulus generation methods, we began preparing to collect data. Before acquiring test subjects, we had to identify our independent, dependent, and control variables. The primary independent variable was the stimulus generation method. We also examined the target sounds provided by the American Tinnitus Association (ATA) and chose two: buzzing and roaring. These two audio files were chosen as target sounds because they represent considerably different spectrums of tinnitus manifestation. Buzzing has high-frequency energy, whereas roaring has low-frequency energy. In more descriptive perceptual terms, buzzing is brighter, while roaring is darker. A more representative sample of the psychoacoustic tinnitus spectrum (PTS) allows us to garner a more encompassing idea of an experimental program's effectiveness. The experimental variables are summarized in Figure 10 below.

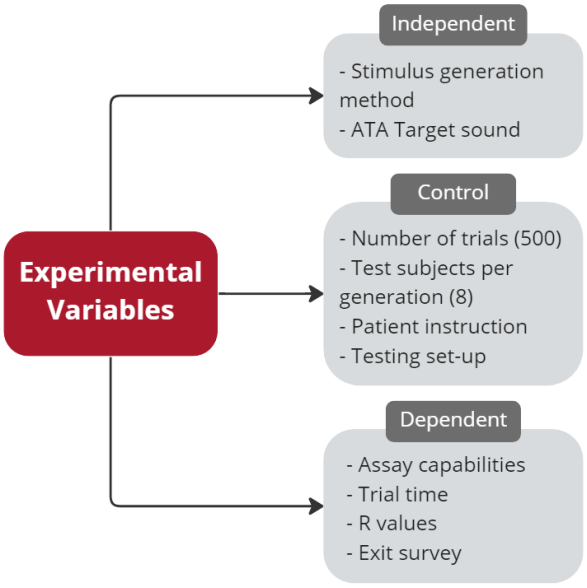


Figure 10: Experimental independent, control, and dependent variables.

Our dependent interest was broadly how well each stimulus generation method could accurately reproduce the target signal. The extent to which each trial result could recreate the target sound was indicated through direct measures such as R-values, indirect measures such as the total time it took to take the test, and the participant's exit survey responses.

3.3.1 Controlled Parameters

The control variables in this experiment were the number of trials (500), the number of test subjects collected for each stimulus generation method (8), patient instruction, and the general testing set-up. It was important to standardize these control variables across data collection trials.

To establish our controls, we first set the amount of data collected from each participant as 500 trials. This number was chosen based on the estimated testing time of 30-40 minutes. We wanted to stay within 1 hour of testing for participant convenience to simplify subject compensation ($10/ hr), and increase people's willingness to take the test repeatedly.

The G*Power application was then used to decide how large populations needed to be to claim statistical significance with 95% certainty (Faul et al, 2007). We assumed a one-tailed t-test with an effect size of two, an α error probability of 0.05, and a power of 0.8. The graphical output of this analysis can be seen in Figure 11 below. Based on this, it was decided that four buzzing roaring subjects would be collected for each stimulus generation method, making eight total test subjects.

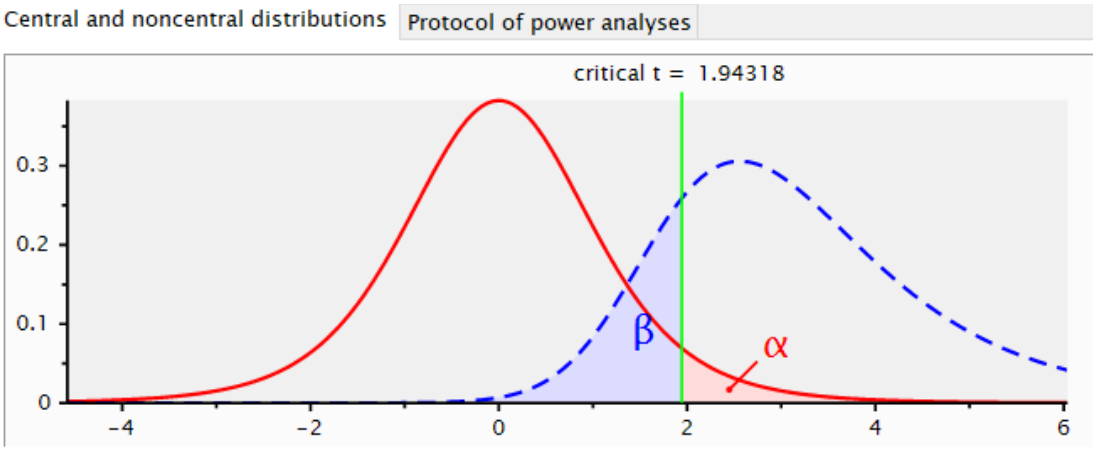


Figure 11: G*Power analysis distribution used to decide sample sizes (Faul et al, 2007).

Next, we standardized the patient instructions. Typically, reverse correlation experiments– such as the visual "S" reconstruction by Gosselin and Schyns (2003)– tell test subjects that their end goal is known to be present in a particular proportion of stimuli. This experiment preserved this idea by telling subjects that their target signal frequency spectra were known to be contained within a certain proportion of stimuli. This is different, however, in the case of actual tinnitus patients because they are the only ones who experience their PTS. If believed relevant to participant confidence in the test, this shortcoming may be combated by explaining that stimuli contain "known tinnitus" frequency spectra. This area could be further explored in the future, but for this research, we focused on the program's ability to produce an accurate tinnitus reconstruction. An example script similar to how we would instruct subjects on how to take the test can be found below:

> "This is an experiment aimed at developing a tinnitus characterization assay. One of the most effective treatments for tinnitus is sound therapy. Similar to exposure therapy, this involves repeatedly playing an external sound for a patient so that their brain will learn to ignore it. The closer the therapy sound is to the patient's tinnitus, the more effective the sound therapy is.
>
> Our experiment specifically tests stimulus generation methods to reconstruct a patient's tinnitus experience. You will hear two sounds: the first is the *target signal* you are trying to reproduce, and the second is a *generated stimulus*. Your job is to determine if the *generated stimuli* have the *target signal* embedded within them. The *generated stimuli* will never sound exactly like the *target signal* because they all have some degree of random noise overlay.
>
> This experiment will likely take about 30 minutes, but you are free to stop or take a break at any time. Your identity will remain anonymous. Do you have any questions?"

The general testing set-up was kept constant amongst all of the buzzing and roaring trials, as this was divided among our group members. In other words, all trials that used a buzzing target signal were conducted on the same laptop. The program's user interface also remained consistent throughout this experiment, with only small aesthetic changes being made for ease of use (see Section 3.2.2 for user interface images).

Controlling for as many variables as possible provided the opportunity to understand better how the stimulus generation method alone affects the efficacy of the tinnitus characterization assay.

3.3.2 Conducting Trials

After identifying the control variables, we began collecting data from test subjects. Test subjects were recruited through friends, acquaintances, and peers. After identifying a test subject, we explained the experiment and obtained their consent using an approved Institutional Review Board (IRB) Consent Form. This form was approved by the UMass Chan Medical School and recognized by a WPI reliance agreement. It can be found in Appendix 9.2. A test subject's identity was then recorded in a 'Subjects Collected Key' to ensure their data was anonymous during analysis. Demographic information such as age, sex, and race were also recorded for participants. A config file was created for each test subject with a unique subject ID and the stimuli generation method of interest.

While the experiment was running, subjects were timed on how long they took to complete five blocks of 100 trials, totaling 500 trials. After each 100 trial block, the subjects were encouraged to take a break if needed. The subjects' observations or comments about the test were also noted to understand how people approach audiologic tasks and auditory stimuli in general. After the experiment, subjects were asked questions from an exit survey, and their answers were recorded using a Likert scale. The questions from this exit survey can be found in Table 2 below.

Table 2: Exit survey questions and Likert scale answers.

Question		1	2	3	4	5
A	I felt like I was able to consistently identify stimuli that had the target signal embedded in them.	Strongly disagree	Moderately disagree	Neither agree nor disagree	Moderately agree	Strongly agree
B	The stimuli sounded natural and like they were being randomly generated.	Strongly disagree	Moderately disagree	Neither agree nor disagree	Moderately agree	Strongly agree

C	The resynthesized version of the noise sounds like the target signal.	Strongly disagree	Moderately disagree	Neither agree nor disagree	Moderately agree	Strongly agree
D	This testing paradigm made sense, was easy to understand and follow.	Strongly disagree	Moderately disagree	Neither agree nor disagree	Moderately agree	Strongly agree

3.3.3 Simulus Generation Methods

When choosing the stimulus generation method, we first began with the Uniform Prior Stimulus Generation method. This method has 100 bins and randomly chooses 30 with a power level of 0 dB, while the rest receive -10 dB. Chosen bins all get the same flat power level, meaning that when the frequency spectrum is played, all frequencies in that bin will be played at the same decibel level. This method was chosen first because it was the standard commonly used in the preliminary testing of the experimental protocol. As a result, it would be an excellent place to start to serve as a benchmark to compare subsequent stimulus generation methods.

All four stimulus generation methods from which data were collected can be found in Table 3 below. These methods were tested chronologically from A to D. The changes from generation to generation are indicated in red.

Table 3: Stimulus Generation Methods A-D evaluated in this study.

Gen	Name	Bin Fill Method	# of Bins	Bin Range	Bin Power	Power Shape
A	Uniform Prior Stimulus Generation	Randomly chosen/ 1 level	100	30	Binary power	Flat
B	Uniform Prior 8 Bin Stimulus Generation	Randomly chosen/ 1 power level	8	3-7	Binary power	Flat
C	Brimijoin Stimulus Generation	Systematic random power choice/ variable power	8	3-7	6 possible powers	Flat
D	Brimijoin Gaussian Smoothed Stimulus Generation	Systematic random power choice	8	3-7	6 possible powers	Gaussian

After utilizing Uniform Prior as Stimulus Generation A, Uniform Prior was used for Stimulus Generation B but with eight bins instead of 100. As such, the minimum number of bins with power was set as three and the maximum as seven to allow for increased variation. This change in bin parameters was made because we believed this would considerably narrow down the possible reconstruction permutations while having a minimal effect on the stimuli from the user's perception. It was also believed that this decreased number of bins might allow for an equal reconstruction accuracy but from fewer trials. This may be something to consider from a

clinical standpoint, as it would shorten the test time for tinnitus patients when creating a reconstructed sound.

The following stimulus generation method tested, Stimulus Generation C, utilized a different method for assigning bin powers. In all generations prior, the power was assigned in a binary fashion of 'high power' or 'low power,' meaning that all bins with power had an equal, "flat" decibel level. On the other hand, this stimulus-generation method was inspired by a 2013 study by Brimijoin et al. Their research used reverse correlation to reconstruct subjects' internal representations of different vowel sounds. However, instead of a binary power value, they set six possible power levels, ranging from -20 dB to 0 dB. As a result, our "Brimijoin" Stimulus Generation had six possible decibel power levels that were randomly chosen for each bin. This was thought to allow for more variation among stimulus generation without having more bins. An example of what frequency spectra from this stimulus generation method might look like can be seen in Figure 12 below.

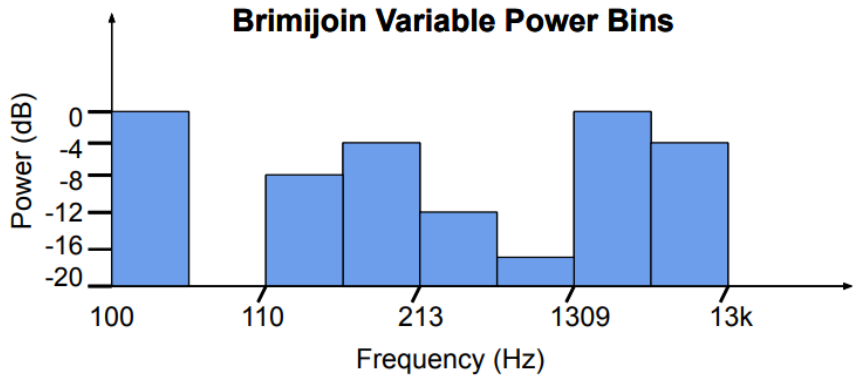


Figure 12: Example Brimijoin frequency bin spectrum.

The final stimulus generation method that was tested– Stimulus Generation D or Brimijoin Gaussian Smoothed– sought to explore the effect of bin shape. Until this point, all bin power was "flat," meaning all frequencies within the bounds of a bin had equal power. Conversely, this stimulus generation method replaced the frequencies within a bin with a Gaussian power distribution level. A representation of this type of frequency spectra can be seen in Figure 13 below.

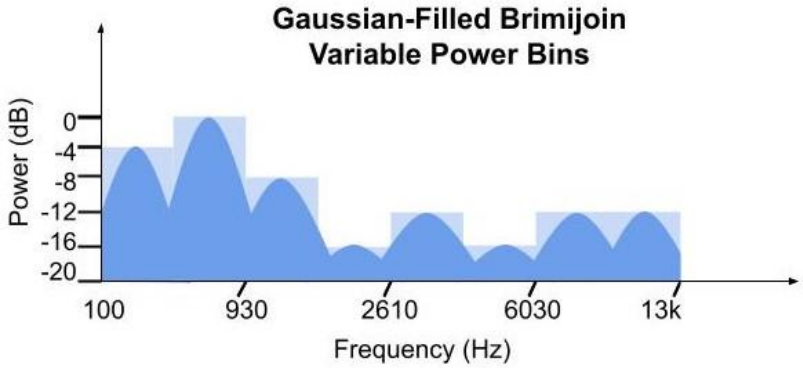


Figure 13: Example Gaussian bin frequency spectrum.

3.3.4 Data Compilation and Analysis

The analysis process after each iteration involved examining the time it took for subjects to complete 500 trials, comparing exit survey results, and primarily calculating the R-values from the subject's reconstruction based on the 'perfect' target signal. R-values based on linear regression and compressive sensing were examined, as opposed to R^2 values, in order to preserve negative correlation coefficients. This was useful because if a subject had a high R^2 value but a negative R-value, it might indicate that they just switched the 'Yes' and 'No' key indicators.

Additionally, once all of the stimulus generation methods were evaluated, an ANOVA test and a Kruskal Wallis test were carried out to determine the statistical significance of differences in the mean R-values across individuals as grouped by "generation," including interaction effects. For our purposes, a one-way ANOVA test was more applicable as we were interested in one dependent variable; the stimulus generation method. A Fisher transformation was performed on this data before the ANOVA test to allow the data to behave as though it is unbounded, even though the R-values are between 0-1. This was necessary because the ANOVA test assumes unbounded data. The ANOVA test also assumes a normal distribution of residuals and is based on a linear regression model. The Kruskal-Wallis test also allows for multiple groups to be compared, but unlike the ANOVA test, it does not assume a normal distribution of residuals, and it does not assume an underlying model type. The Kruskal-Wallis is a nonparametric test and, as a result, is more robust than the ANOVA test.

The MATLAB script used to perform the ANOVA statistical analysis can be found in Appendix 9.5. The Kruskal Wallis statistical analysis script was identical to this, except it used the "[p, tbl, stats] = kruskalwallis()" function instead of the "[p, tbl, stats] = anova1()" function and the Fisher ("atanh()") transformation was not performed. Post-hoc tests were also conducted to determine pairwise differences in mean R-value between generations using the "multcompare()" function.

In broad terms, these statistical tests allowed us to test the null hypothesis. The null hypothesis is the default idea that no statistically significant difference exists in the mean of two

or more data sets. In our case, “data sets” are R-values of subject trials performed using stimulus generations A through D. If the p-value that results from a statistical significance test is less than 0.05, then the null hypothesis is *rejected*, meaning that there *is* a statistical significant difference in data set means.

3.3.5 Iterative Approach to Stimulus Generation Methods

We wanted to ensure that, throughout our experiment, we had the chance to choose our stimulus generation methods iteratively. In other words, we could not choose our next stimulus generation method– after Uniform Prior– because we wanted it to be dynamically informed by the data as we collected it. We decided on a strict schedule of roughly 2-week iterations to facilitate this. We collected data and analyzed our results to choose a stimulus-generation method for the next iteration. Table 4 below displays our general plan and schedule for these iterations.

Table 4: General iterative project timeline (as a Gantt Chart).

Num.	Type	Task	A Term							B Term							C Term							D Term						
			1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	Resrch	Understanding topic																												
2	Resrch	Literature Reveiw																												
3	Dev	Defining Design Space																												
4	Dev	Test pre-clinincal Assay																												
5	Dev	Proposing MQP																												
6	Dev	Problem Statement																												
7	Resrch	Prepared for SGA																												
8	Itor	Collect SGA Data																												
9	Itor	Collect SGB Data																												
10	Dev	Analyze/ Iterate SGA and B																												
11	Itor	Collect SGC Data																												
12	Dev	Analyze/ Iterate SGC																												
13	Dev	Create SGD																												
14	Itor	Collect SGD Data																												
15	Dev	Analyze SGD																												
16	Reprt	Draft Paper																												
17	Misc	Presentation Prep																												

4. Results

The test subjects that performed trials were recorded in a key to ensure anonymity. This key also ensured they signed the consent form and noted their demographic information. The information recorded was: age, gender (M or F), race (white, black, Asian, American Indian, Pacific Islander, or more than one race), and if they were of Hispanic/Latino descent. Table 5 below shows this key without the identity of the test subjects. Moreover, the stimulus generation method, target sound, date, and proctor were also noted.

Table 5: Test subject demographic information.

Proctor	ID	Age	Gender	Race	Hispanic/ Latino	Stim Gen	Target Sound
M	Subject 1	21	F	white	N	A	Buzzing
	Subject 2	22	M	asian	N		
	Subject 3	21	F	white	N		
	Subject 4	52	M	white	N		
J	Subject 5	21	M	white	N		Roaring
	Subject 6	22	M	white	N		
	Subject 7	20	M	white	N		
	Subject 8	20	M	white	N		
M	Subject 1	21	F	white	N	B	Buzzing
	Subject 9	52	F	white	N		
	Subject 4	52	M	white	N		
	Subject 10	21	F	white	N		
J	Subject 5	21	M	white	N		Roaring
	Subject 11	21	F	white	N		
	Subject 12	21	M	white	N		
	Subject 13	18	M	white	N		
M	Subject 1	21	F	white	N	C	Buzzing
	Subject 2	22	M	asian	N		
	Subject 14	21	F	asian	N		
	Subject 10	21	F	white	N		
J	Subject 15	21	M	white	N		Roaring
	Subject 5	21	M	white	N		
	Subject 16	21	F	white	N		
	Subject 17	21	M	white	N		
M	Subject 1	21	F	white	N	D	Buzzing
	Subject 18	21	F	white	N		
	Subject 2	22	M	asian	N		
	Subject 10	21	F	white	N		
J	Subject 5	21	M	white	N		Roaring
	Subject 6	22	M	white	N		
	Subject 19	21	F	white	N		
	Subject 20	21	F	white	N		

The average age of all test subjects was 24 years old. Additionally, 15 test subjects were female, and 17 were male. This equates to a 46.9% female and 53.1% male test population. Of

the subjects tested, 87.5% were white, and 12.5% were Asian. No test subjects reported being of Hispanic or Latino descent.

Quantitative information on test subjects was also collected. While completing 500 trials, test subjects were timed to compare the efficiency of the different stimulus-generation methods. The R correlation coefficients between the subject reconstruction and the actual target signal were also recorded to assess stimuli generation accuracy. R-values, along with their corresponding R^2 values, based on a linear reconstruction (LR) and a compressive sensing (CS) approach, for each test subject can be found in Appendix 9.3 at the end of this report. The R-values from the linear regressions of all four stimulus generation methods examined can be compared in the box and whisker plot in Figure 14 below.

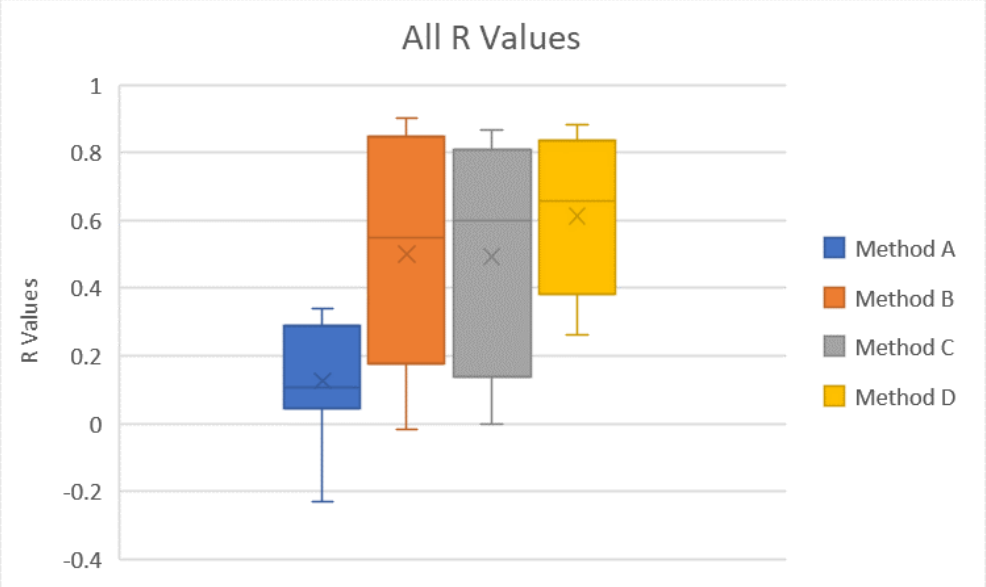


Figure 14: All stimulus generation methods R-value box and whisker plot.

In addition to these quantitative values, qualitative exit survey results were recorded from many test subjects after their trials. These results can be found in Table 6 below.

Table 6: Test subject exit survey results.

ID	Stim Gen	Sound	Q1	Q2	Q3	Q4
Subject 1	A	Buzz	2	2	---	4
Subject 2	A	Buzz	1	3	---	4
Subject 3	A	Buzz	---	---	---	---
Subject 4	A	Buzz	1	1	---	3
Subject 5	A	Roar	---	---	---	---
Subject 6	A	Roar	---	---	---	---
Subject 7	A	Roar	---	---	---	---

Subject 8	A	Roar	---	---	---	---
		Average	1.3	2.0	---	3.7
Subject 1	B	Buzz	4	4	---	4
Subject 9	B	Buzz	2	1	---	2
Subject 4	B	Buzz	2	2	---	3
Subject 10	B	Buzz	4	4	---	5
Subject 5	B	Roar	---	---	---	---
Subject 11	B	Roar	---	---	---	---
Subject 12	B	Roar	---	---	---	---
Subject 13	B	Roar	---	---	---	---
		Average	3.0	2.8	---	3.5
Subject 1	C	Buzz	3	4	---	4
Subject 2	C	Buzz	4	3	---	4
Subject 14	C	Buzz	2	5	---	4
Subject 10	C	Buzz	4	4	---	5
Subject 15	C	Roar	3	1	---	4
Subject 5	C	Roar	4	3	---	5
Subject 16	C	Roar	3	1	---	4
Subject 17	C	Roar	4	3	---	5
		Average	3.4	3.0	---	4.4
Subject 1	D	Buzz	4	4	--	4
Subject 18	D	Buzz	4	4	--	5
Subject 2	D	Buzz	3	5	--	4
Subject 10	D	Buzz	4	4	--	5
Subject 5	D	Roar	4	3	--	5
Subject 6	D	Roar	2	4	---	4
Subject 19	D	Roar	---	---	---	---
Subject 20	D	Roar	---	---	---	---
		Average	3.5	4.0	---	4.5

4.1 Stimulus Generation A: Uniform Prior

Uniform Prior was chosen as stimulus generation A because it was previously believed to be the most reliable. It was used as the “standard” protocol in preliminary testing of the MATLAB program and general experimental paradigm. The demographic information of the eight subjects from both the buzzing and roaring trials can be found in Table 5 above. From this

stimulus generation method, 6/8 of test subjects were male, 7/8 were white,and the remainder were Asian. Additionally, as seen in Appendix 9.3, the average time it took subjects to complete 500 trials was around 33.5 ± 12.5 minutes.

A graphical representation of all bins from each subject of the buzzing trials can be found in Figures 15 and 16 below. In this figure, black is the true target signal. The dashed lines are the linear reconstructions, whereas the solid lines are the compressive sensing reconstructions.

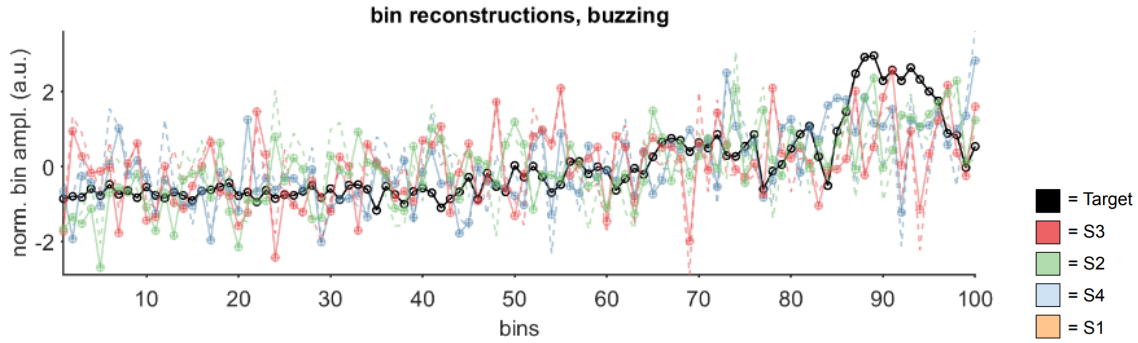


Figure 15: Uniform Prior stimulus generation buzzing subject response bin reconstruction.

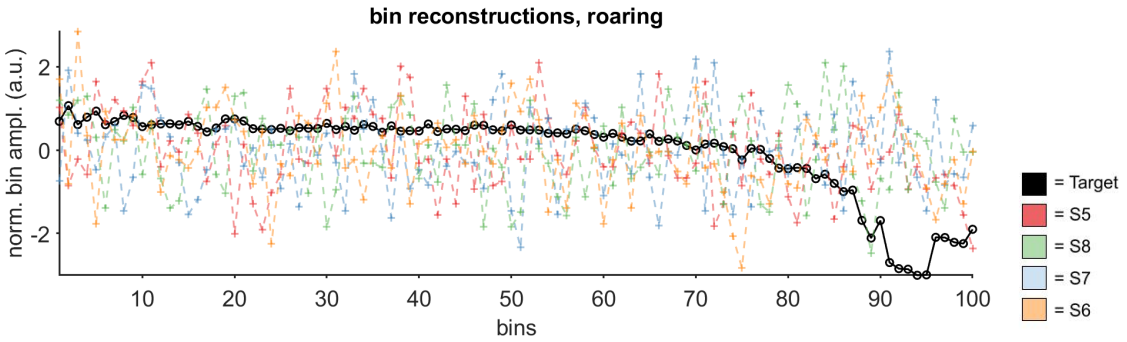


Figure 16: Uniform Prior stimulus generation roaring subject response bin reconstruction.

4.1.1 Quantitative

The R-values for the Uniform Prior generation method were calculated using both linear regression and compressive sensing methodologies, listed in Appendix 9.3. Using linear regression, Stimulus Generation A had an average R-value of 0.126 ± 0.183, meaning the R^2 value was 0.045 ± 0.043. On the other hand, the compressive sensing R-value was 0.205 ± 0.314, so the R^2 value was 0.128 ± 0.133. The distribution of linear regression R-values is shown in the Figure 17 box and whisker plot below.

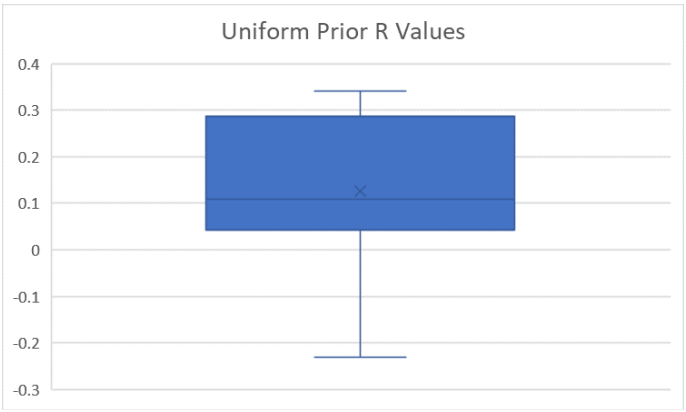


Figure 17: Uniform Prior R-values box and whisker plot.

4.1.2 Qualitative

Test subjects who took the Uniform Prior generation method were asked Question 1, Question 2, and Question 4 of the exit survey. They were asked to answer on a scale from 1 to 5, with 5 being "strongly agree" and 1 being "strongly disagree." In response to Question 1, "I felt like I was able to consistently identify stimuli that had the target signal embedded in them," subjects responded 1.3 on average. In response to Question 2, "The stimuli sounded natural and like they were being randomly generated," subjects responded 2.0 on average. In response to Question 4, "This testing paradigm made sense, was easy to understand and follow," subjects responded 3.7 on average.

Other qualitative feedback from this stimulus generation method came as comments from subjects while taking the MATLAB protocol. Many comments were made about how 'all of the stimuli sounded the same' and 'they all sounded like white noise.' Subjects also reported that the trials felt like they took a long time.

In addition to receiving low analytical scores, this stimulus-generation method was also poorly received by the test subjects. Test subjects reported an inability to consistently identify stimuli with the control sound within them. Many comments were made about how all of the stimuli sounded the same and that they all sounded like white noise. This was also reflected numerically, as one of the exit survey questions averaged near a 1, the lowest score possible. Furthermore, the second exit survey question had an average score of 2 for this section, further showing that the test subjects did not feel like the stimuli were being randomly generated.

4.2 Stimulus Generation B: Uniform Prior 8 Bin

Uniform Prior 8 with eight bins instead of 100 was chosen for Stimulus Generation B. The demographic information of the eight subjects from both the buzzing and roaring trials can be found in Table 5 above; the ratio of male to female test subjects was 50:50, and all subjects were white. As seen in Appendix 9.3, the average time it took subjects to complete 500 trials was around 30.0 ± 6.0 minutes.

A graphical representation of all bins from each subject of the buzzing trials can be found in Figures 18 and 19 below. In this figure, the eight bins are substantially easier to differentiate when compared to the 100 bins of Stimulus Generation A. Black is the true target signal; the dashed lines are the linear reconstructions; the solid lines are the compressive sensing reconstructions. Because only eight bins were used in this generation, the resulting compressive sensing values were identical, so some solid lines overlapped.

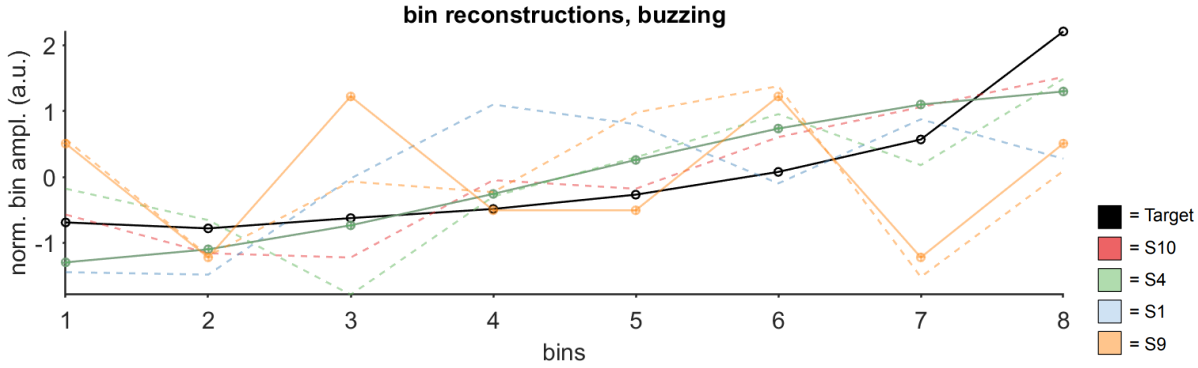


Figure 18: Uniform Prior 8 Bin stimulus generation buzzing subject response bin reconstruction.

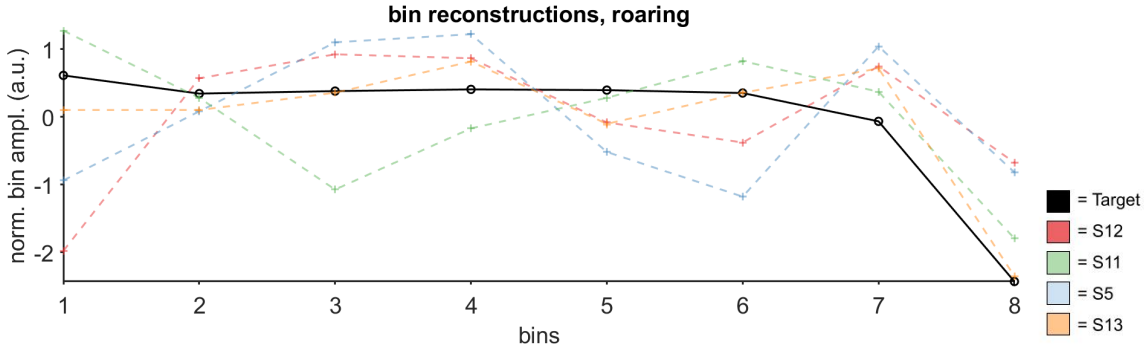


Figure 19: Uniform Prior 8 Bin stimulus generation roaring subject response bin reconstruction.

4.2.1 Quantitative

The R-values for the Uniform Prior 8 bin generation method were calculated using both linear regression and compressive sensing methodologies, listed in Appendix 9.3. Using linear regression, this stimulus generation had an average R-value of 0.501 ± 0.358, meaning the R^2 value was 0.363 ± 0.346. On the other hand, the compressive sensing R-value was 0.531 ± 0.279, so the R^2 value was 0.350 ± 0.293. The distribution of linear regression R-values is shown in the Figure 20 box and whisker plot below.

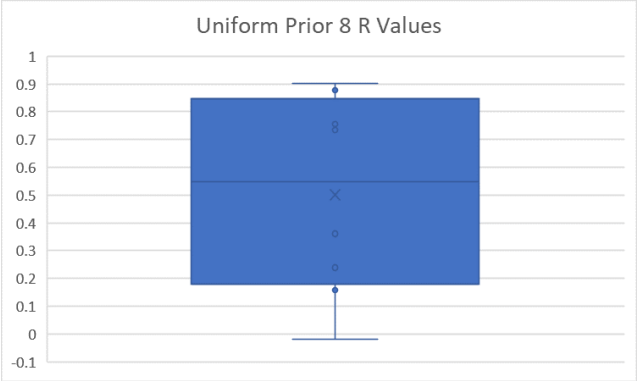


Figure 20: Uniform Prior 8 Bin R-values box and whisker plot.

4.2.2 Qualitative

Test subjects who took the Uniform Prior 8 bin generation method were asked Question 1, Question 2, and Question 4 of the exit survey. They were asked to answer on a scale from 1 to 5, with 5 being "strongly agree" and 1 being "strongly disagree." In response to Question 1, "I felt like I was able to consistently identify stimuli that had the target signal embedded in them," subjects responded 3.0 on average. In response to Question 2, "The stimuli sounded natural and like they were being randomly generated," subjects responded 2.8 on average. In response to Question 4, "This testing paradigm made sense, was easy to understand and follow," subjects responded 3.5 on average.

Other qualitative feedback from this stimulus generation method came as comments from subjects while taking the MATLAB protocol. Many comments were made about how the protocol was 'hard' and that trials feel more difficult as they go on. Subjects also reported that the trials felt like they took a long time.

This iteration showed improved distinguishability as the subjects reported being able to discern more easily between the presented stimuli. This was also reflected in the exit survey results, which showed that subjects no longer "strongly disagreed" with the statement that they could consistently identify stimuli the target signal embedded within them. Instead, the average response to this question was a 3.0, which indicates "neither agree nor disagree." The second question also had an improved score, which indicated that the test subjects could more easily discern the difference between the stimuli presented to them.

4.3 Stimulus Generation C: Brimijoin

The Brimijoin method was chosen for Stimulus Generation C to add variation among possible power levels with the newly decreased number of bins. The demographic information of the eight subjects from both the buzzing and roaring trials can be found in Table 5 above; the ratio of male to female test subjects was 50:50. Additionally, 2/8 subjects were Asian, with the rest being white. As seen in Appendix 9.3, the average time it took the subjects to complete 500 trials was around 24.8 ± 6.1 minutes.

A graphical representation of all bins from each subject of the buzzing trials can be found in Figures 21 and 22 below. The black is the true target signal, and the dashed lines are linear reconstructions, while the solid lines are compressive sensing reconstructions. Once again, because only eight bins were used, the compressive sensing values can be similarly causing some of the solid lines to overlap.

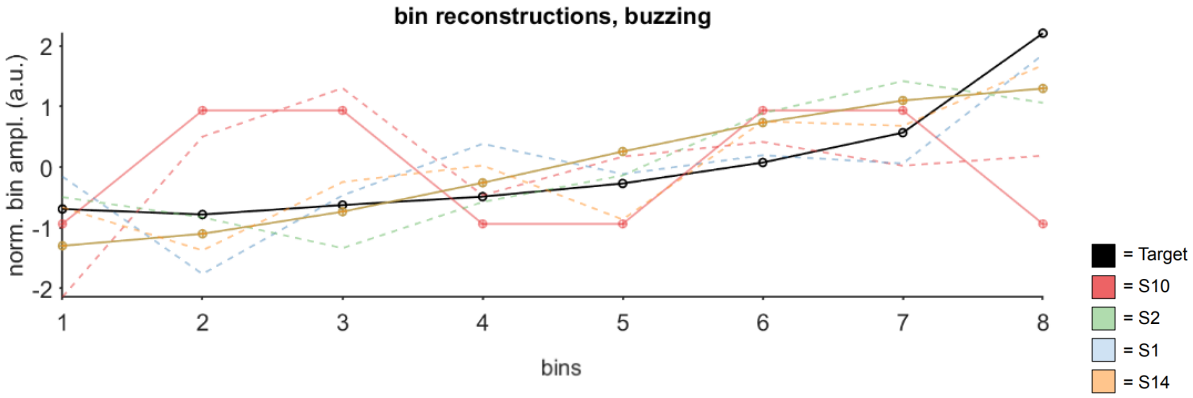


Figure 21: Brimijoin stimulus generation buzzing subject response bin reconstruction.

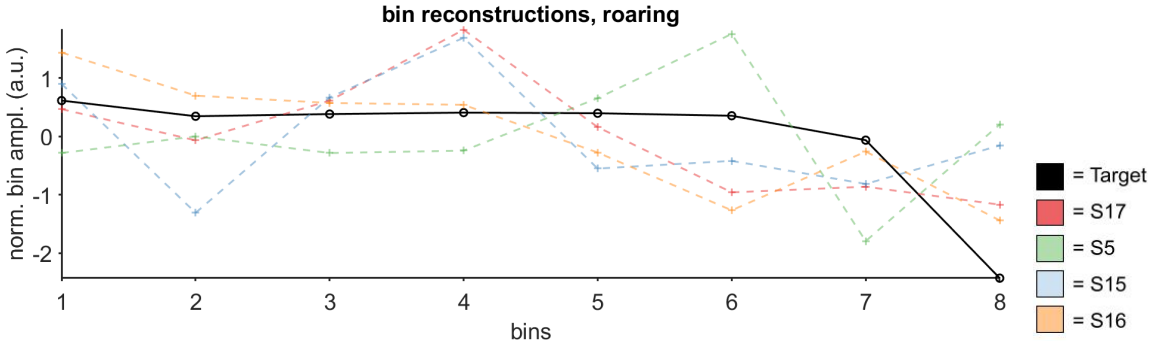


Figure 22: Brimijoin stimulus generation roaring subject response bin reconstruction.

4.3.1 Quantitative

The R-values for the Brimijoin stimulus generation method were calculated using linear regression and compressive sensing methodologies, listed in Appendix 9.3. Using linear regression, this stimulus generation had an average R-value of 0.493 ± 0.346, meaning the R^2 value was 0.348 ± 0.310. On the other hand, the compressive sensing R-value was 0.506 ± 0.376, so the R^2 value was 0.380 ± 0.295. The distribution of linear regression R-values is shown in the Figure 23 box and whisker plot below.

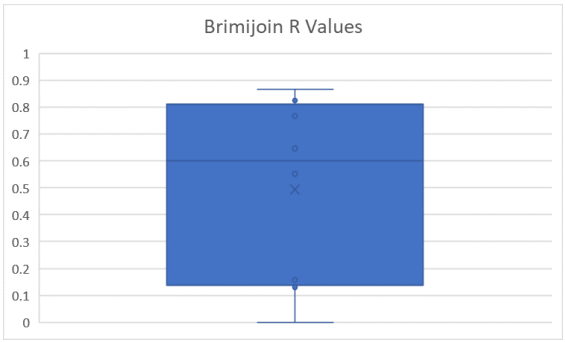


Figure 23: Brimijoin R-values box and whisker plot.

4.3.2 Qualitative

Test subjects who took the Brimijoin generation method were asked Question 1, Question 2, and Question 4 of the exit survey. They were asked to answer on a scale from 1 to 5, with 5 being "strongly agree" and 1 being "strongly disagree." In response to Question 1, "I felt like I was able to consistently identify stimuli that had the target signal embedded in them," subjects responded 3.4 on average. In response to Question 2, "The stimuli sounded natural and like they were being randomly generated," subjects responded 3.0 on average. In response to Question 4, "This testing paradigm made sense, was easy to understand and follow," subjects responded 4.4 on average.

Other qualitative feedback from this stimulus generation method came as comments from subjects while taking the MATLAB protocol. Many comments were made about how the protocol feels more difficult as they go on. Subjects also reported that the trials felt like they took a long time. These comments, however, were less frequent as more subjects were repeat subjects.

Another benefit of the Brimijoin testing method was indicated in the exit survey. This is the first generation method that produced a positive average answer to the first question. In this instance, the test subjects were more confident that they could identify stimuli with the target signal within them.

4.4 Stimulus Generation D: Brimijoin Gaussian Smoothed

The Brimijoin method was modified to become the Brimijoin Gaussian Smoothed method for Stimulus Generation D. This method had Gaussian bin powers instead of flat bin powers to make stimuli sound more natural. The demographic information of the eight subjects from both the buzzing and roaring trials can be found in Table 5 above. Among these test subjects, 5/8 were female, and 3/5 were male. Additionally, 1/8 subjects were Asian, with the rest being white. As seen in Appendix 9.3, the average time it took the subjects to complete 500 trials was around 28.6 ± 4.6 minutes.

A graphical representation of all bins from each subject of the buzzing trials can be found in Figures 24 and 25 below. Please note that due to the nature of the reconstruction code of the tinnitus assy, the black target signal bin power assumes a level, flat power level in each bin. As a

result, these graphs are not showing the exact target signal, but the R-values of Appendix 9.3 still offer an accurate comparison. Additionally, the dashed lines are linear reconstructions, while the solid lines are automatically generated compressive sensing reconstructions by the assay. As mentioned, our analysis did not consider these because of their decreased sensitivity with fewer bins.

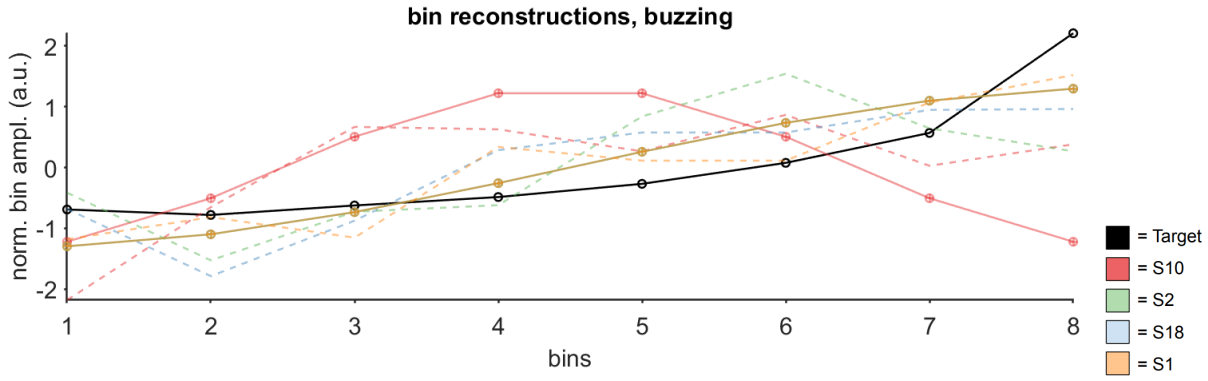


Figure 24: Brimijoin Gaussian Smoothed stimulus generation buzzing subject response bin reconstruction.

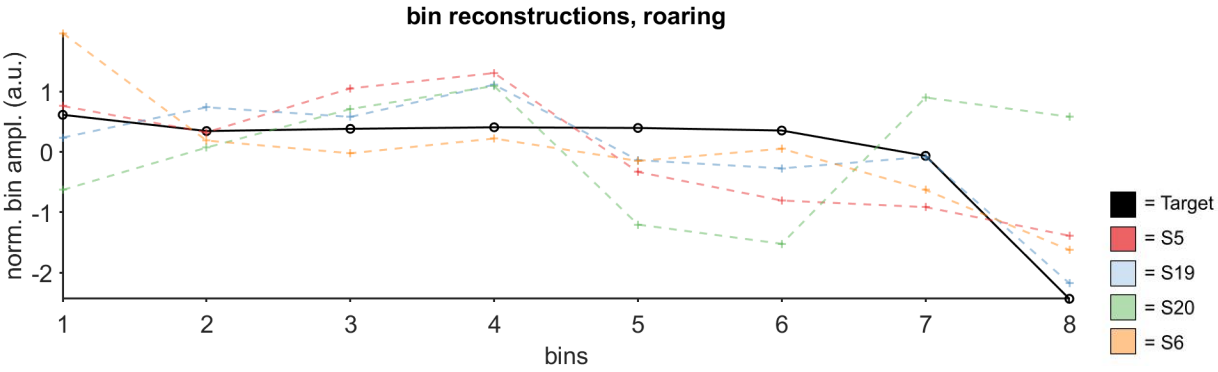


Figure 25: Brimijoin Gaussian Smoothed stimulus generation roaring subject response bin reconstruction.

This stimulus generation method differed from all others because the target signals could not be approximated with flat bins, but they had to be recreated with Gaussian approximations. These approximations can be seen in Figures 26 and 27 below. The code written to generate these approximations is also in Appendix 9.4.

Buzzing Target Signal	**Roaring Target Signal**

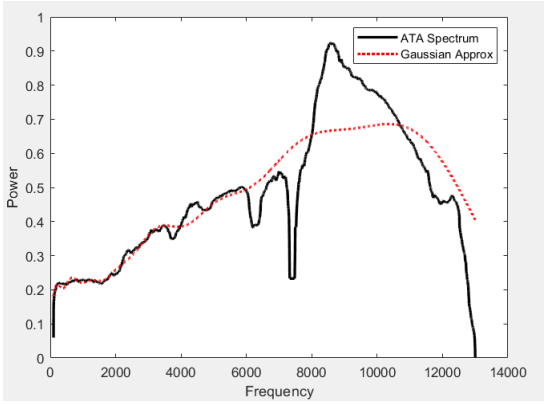


Figure 26: Buzzing Gaussian approximation.

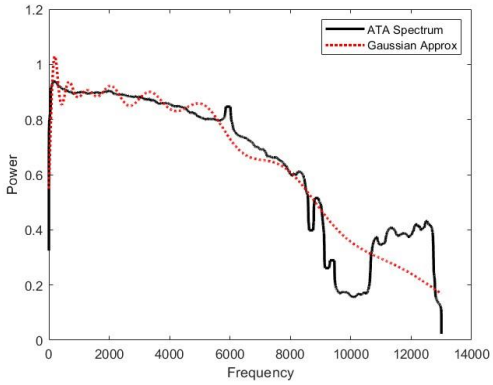


Figure 27: Roaring Gaussian approximation.

4.4.1 Quantitative

The R-values for the Brimijoin Gaussian Smoothed stimulus generation method were calculated through only linear regression because compressive sensing had proven to be unreliable in the cases of a smaller bin number. Nevertheless, linear regression results can be found in Appendix 9.3. Using linear regression, this stimulus generation had an average R-value of 0.615 ± 0.240, meaning the R^2 value was 0.427 ± 0.282. The distribution of linear regression R-values is shown in the Figure 28 box and whisker plot below.

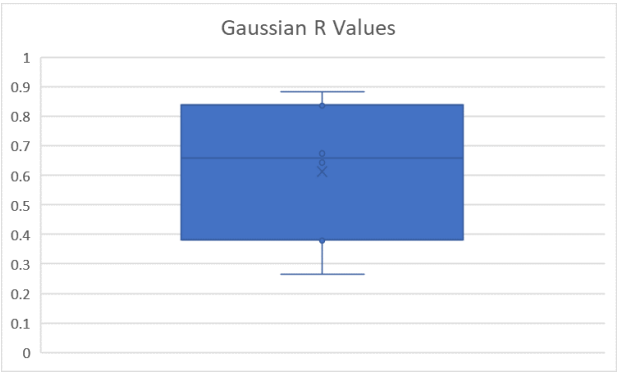


Figure 28: Brimijoin Gaussian Smoothed R-values box and whisker plot.

4.4.2 Qualitative

Test subjects who took the Brimijoin generation method were asked Question 1, Question 2, and Question 4 of the exit survey. They were asked to answer on a scale from 1 to 5, with 5 being “strongly agree” and 1 being “strongly disagree.” In response to Question 1, “I felt like I was able to consistently identify stimuli that had the target signal embedded in them,” subjects responded 3.5 on average. In response to Question 2, “The stimuli sounded natural and like they were being randomly generated,” subjects responded 4.0 on average. In response to Question 4,

“This testing paradigm made sense, was easy to understand and follow,” subjects responded 4.5 on average.

Other qualitative feedback from this stimulus generation method came as comments from subjects while taking the MATLAB protocol. Like other stimulus generation methods, subjects commented that the trials took a long time. These comments, however, were less frequent as more subjects were repeat subjects. Additionally, subjects expressed worry about ‘getting a good grade.’

4.5 Statistical Significance and Correlation Analysis

When performing statistical significance testing, several different comparison methods were chosen to construct a complete picture of how the different dependent variables may be related. Both linear reconstruction and compressive sensing R-values were considered for the ANOVA and the Kruskal Wallis analysis; however, compressive sensing values were not calculated for Stimulus Generation D, the Brimijoin Gaussian Smoothed. This is because compressive sensing capabilities drastically decrease with fewer bins. This was evident as the compressive sensing R-values would return identical results for different subjects that linear regression R-values differentiated between. The five relationships examined were as follows:

1. Only the buzzing trials from all four stimulus generation methods (four groups for linear reconstruction, four for compressive sensing)
2. Only the roaring trials from all four stimulus generation methods (four groups for linear reconstruction, three for compressive sensing)
3. All of the buzzing trials vs. all of the roaring trials (two groups)
4. All of the buzzing trials and all of the roaring trials from all four stimulus generation methods (four groups for linear reconstruction, three for compressive sensing)
5. All of the buzzing trials vs. all of the roaring trials from all four stimulus generation methods (eight groups for linear reconstruction, six for compressive sensing)

The resulting P values of all 20 of these combinations can be seen in Table 7 below. Two comparisons that a P value < 0.05, meaning that the null hypothesis could be rejected and that there was a statistically significant difference between some groups. The trials found to have statistically significant differences were the ANOVA and Kruskal Wallis tests on the linear regression R-values of all buzzing and roaring trials from all four stimulus generation methods (four groups for linear regression).

Table 7: P values of ANOVA and Kruskal Wallis statistical significance tests.

	P Values			
	ANOVA		Kruskal Wallis	
Comparison (# of groups)	LR	CS	LR	CS
A-D only buzzing (4, 3)	0.1828	0.3029	0.2065	0.1923
A-D only roaring (4, 3)	0.0799	0.0898	0.0685	0.0825
Buzzing vs. roaring (2)	0.7882	0.1031	0.7344	0.1029
A-D (4, 3)	0.0256	0.0885	0.0239	0.1100
A-D buzz vs. roar (8, 6)	0.0788	0.1454	0.1021	0.2105

The statistically significant ANOVA box and whisker plot can be seen in Figure 14 above. Additionally, the test plots can be found in Appendix 9.7 and 9.8 at the end of this report. In addition to each of these box and whisker plots, post hoc tests were also conducted to determine pairwise differences in mean R-values amongst test groups. These can also be found in Appendix 9.7 and 9.8. The pairwise comparisons, for both the ANOVA and Kruskal Wallis tests, revealed that it was Stimulus Generation A and D between which there was a statistically significant difference.

5. Discussion

The data summarized above from four stimulus generation methods– Uniform Prior, Uniform Prior 8 Bin, Brimijoin, and Brimijoin Gaussian Smoothed– provides insight into the performance of this tinnitus characterization assay under different design configurations. This data can be examined to understand better how reverse correlation (RC) may be implemented to precisely estimate a patient's constant, non-tonal psychoacoustic tinnitus spectrum (PTS).

5.1 Comparing Stimulus Generation Methods

The Uniform Prior stimulus generation method was chosen for the first method, Stimulus Generation A, because it had been used the most in past testing of the tinnitus assay. This means it served as a benchmark, or control, against which we could compare other stimulus generation methods. The average linear regression R-value was 0.126 ± 0.183 (R^2 value 0.045 ± 0.043), and the compressive sensing R-value was 0.205 ± 0.314 (R^2 value 0.128 ± 0.133). From the exit surveys, we also determined that test subjects generally disagreed with the statements “I felt like I was able to consistently identify stimuli that had the target signal embedded in them,” and “The stimuli sounded natural and like they were being randomly generated.” Subjects somewhat agreed with the statement that “This testing paradigm made sense, was easy to understand and follow.”

For Stimulus Generation B– Uniform Prior 8 Bin– we decreased the number of bins from 100 to 8 to decrease the number of trials needed to get a good sound reconstruction at a minimal expense from the user's perception. The average linear regression R-value was 0.501 ± 0.358 (R^2

value 0.363 ± 0.346), and the compressive sensing R-value was 0.531 ± 0.279 (R^2 value 0.350 ± 0.293). From the exit surveys, we also determined that test subjects were generally neutral on the statement "I felt like I was able to consistently identify stimuli that had the target signal embedded in them," and slightly disagreed with the statement "The stimuli sounded natural and like they were being randomly generated." Subjects somewhat agreed with the statement that "This testing paradigm made sense, was easy to understand and follow."

The average linear regression R-value almost quadrupled from Stimulus Generation A to Stimulus Generation B, indicating that fewer bins improved reconstruction accuracy. This increase was also reflected in compressive sensing R-values, but this method was less sensitive for stimulus generation methods with low numbers of bins. In other words, it would report the same R-values for different subjects where linear regression could differentiate. Stimulus Generation B also had a greater standard deviation, or spread, than Stimulus Generation A by about 8-fold. Regarding qualitative exit survey results, Question 1 and Question 2 had slightly better responses from Stimulus Generation A to Stimulus Generation B, whereas Question 4 slightly decreased.

For Stimulus Generation C– Brimijoin– we utilized a different method for filling frequency bins; instead of a binary "filled" or "not filled" power value, they had six possible power levels. The levels were chosen randomly, and the rationale was to emulate the 2013 study by Brimijoin et al. This change was also thought to bring back some variability lost from decreasing the bin size. The average linear regression R-value was 0.493 ± 0.346 (R^2 value 0.348 ± 0.310), and the compressive sensing R-value was 0.506 ± 0.376 (R^2 value 0.380 ± 0.295). From the exit surveys, we also determined that test subjects generally moderately agreed with the statements "I felt like I was able to consistently identify stimuli that had the target signal embedded in them," and "The stimuli sounded natural and like they were being randomly generated." Subjects strongly agreed with the statement that "This testing paradigm made sense, was easy to understand and follow."

The standard deviation of Stimulus Generation C was about equal to Stimulus Generation B. The average linear regression R-values of Stimulus Generation C were almost quadrupled from Stimulus Generation A but were about equal to those of Stimulus Generation B. This indicates that the increased number of possible bin powers did not substantially improve reconstruction accuracy. R-values of Stimulus Generation C had a standard deviation very similar to Stimulus Generation B, about 8x larger than Stimulus Generation A. Regarding qualitative exit survey results, Question 1, Question 2, and Question 3 all had the highest average responses of any method before Stimulus Generation C.

For Stimulus Generation D– Brimijoin Gaussian Smoothed– we utilized a different shape for filling frequency bins; instead of a flat or constant power value, the frequencies within a bin had a Gaussian power distribution level. This change made the stimuli sound more natural to the human ear. The average linear regression R-value was 0.615 ± 0.240 (R^2 value 0.427 ± 0.282). From the exit surveys, we also determined that test subjects generally moderately agreed with the statement "I felt like I was able to consistently identify stimuli that had the target signal

embedded in them," and strongly agreed with the statement "The stimuli sounded natural and like they were being randomly generated." Subjects strongly agreed with the statement that "This testing paradigm made sense, was easy to understand and follow."

The average linear regression R-values of Stimulus Generation D were the largest of any stimulus generation method, at roughly 5x larger than Stimulus Generation A, 2x larger than Stimulus Generation B, and 2x larger than Stimulus Generation C. This indicates that Stimulus Generation D had the highest accuracy. R-values of Stimulus Generation D had a standard deviation slightly smaller than Stimulus Generation B and C but still larger than Stimulus Generation A. Regarding qualitative exit survey results, Question 1, Question 2, and Question 3 all had the highest overall average responses from Stimulus Generation D.

5.2 The Best Stimulus Generation Method

Revisiting the primary goal of this project, we wanted to improve the efficacy of an assay that can accurately and precisely estimate a patient's psychoacoustic tinnitus spectrum (PTS), specifically a constant, non-tonal PTS. This development is intended to enable more effective sound therapy and therefore create better methods of treating tinnitus heterogeneity. We focused on importing this tinnitus characterization assay by specifically examining stimuli generation methods. Through this experiment, we determined that of the four stimulus generation methods evaluated, Stimulus Generation D: Brimijoin Gaussian Smoothed was the best method for this assay.

The Brimijoin Gaussian Smoothed stimulus generation method was the best in quantitative and qualitative evaluations. This stimulus-generating method had the largest average linear regression R-values and the highest— or most positive— exit survey results. These values can be seen in Appendix 9.3 (see Section 5.1 for a more detailed chronological comparison of stimulus generation methods). Additionally, Stimulus Generation D was found to have a statistically significant difference from Stimulus Generation A via an ANOVA and a Kruskal Wallis statistical significance (and post hoc) test.

5.3 Other Observations

Although the primary focus of this experiment was determining which stimulus generation method was most effective in characterizing, or reproducing, the given target signal. Other observations can be made that are worth discussing further. For example, we recorded the time it took subjects to complete their 500 trials, and this revealed that subjects took on average: 33.5 ± 12.5 minutes for Stimulus Generation A, 30.0 ± 6.0 minutes for Stimulus Generation B, 24.8 ± 6.1 minutes for Stimulus Generation C, and 28.6 ± 4.6 minutes for Stimulus Generation D. Compared to Stimulus Generation A, Stimulus Generation B took users about 3 minutes (10%) less on average to complete 500 trials with half as large of a standard deviation. The next method, Stimulus Generation C, took users the least amount of time to complete 500 trials; about 5 minutes (17%) less than Stimulus Generation B and 8 minutes (24%) less than Stimulus

Generation A. Stimulus Generation D then took users the second-to-least amount of time to complete 500 trials; about 5 minutes (15%) less than Stimulus Generation A, about 1 minute (3%) less than Stimulus Generation B, and about 4 minutes (17%) more than Stimulus Generation C. The standard deviation, however, was the smallest observed of any method.

Moreover, observing test subject comments and attitudes were very interesting throughout this process. We observed people seemingly ready and willing to participate and then some dramatic changes in morale as they began trials. The test protocol was primarily perceived as tedious or challenging, especially as subjects processed through the 500 trials. Asking test subjects to focus on auditory stimuli, instead of visual or tactile stimuli, takes greater relative concentration. We believe this is because we do not necessarily rely as much on sound as we do sight, so subjects feel an increased pressure to perform well.

Another potential factor influencing test subjects' stress is that we, the test proctors, know many of them personally. Asking friends or family members to participate in research, especially related to our Major Qualifying Project, was perceived as a "high stakes" task by many subjects. This performance anxiety could have influenced test results positively or negatively compared to a hypothetical clinical setting.

5.4 Sources of Error

Throughout this experiment, certain aspects had the potential to be influenced by human error or our specific testing circumstances. Many test subjects completed trials for multiple stimulus generation methods. This means that they may have grown more familiar with the testing paradigm as they completed more trials, which could have potentially biased their qualitative exit survey results. This phenomenon is known as the *carryover effect*. For example, a test subject may have found that 'the testing paradigm made more sense, and was easier to understand and follow' upon their second time performing trials after they had some time to process the concept. This would presumably affect their response to Question 4 of the exit survey, which is related to understanding the testing paradigm.

Other sources of error include a lack of strict control of specific testing conditions. For example, subjects used different headphone types and brands and listened to the experimental protocol at different volumes. Additionally, the environment in which test subjects took this assay was highly variable, meaning there was a substantial potential for environmental distractions, such as ambient noises or visual distractions.

Moreover, there are inherently some subjective hearing differences amongst test subjects. None of the subjects specifically had tinnitus, but there could be slight variations in how subjects perceive auditory stimuli. For example, some subjects were 50+ years old, meaning they could experience average age-related hearing loss, potentially affecting their target signal reconstructions.

6. Recommendation

As mentioned previously, the ultimate goal for the overarching research is to develop an assay that can effectively replicate a tinnitus patient's PTS to assist in future therapies or therapy research. To effectively do so, we feel that further research would benefit some areas of the paradigm. The first step in this process is to refine the design space that was previously mentioned. Beyond this, there are alterations to this process that would need to be made to make this paradigm fit for clinical purposes.

6.1 Design Space Alterations

Patient instruction is the first area of the design space that could use some reworking. In this experiment, the instructions needed to be standardized, leading to questions that may not have been necessary. This also permitted different understandings of what the test subject was looking for within the stimuli, which may have altered the results from individual to individual. Some test subjects mentioned after their experiment that they only gave a "yes" answer a handful of times because they were so intent on looking for exact matches to the target sound. On the other hand, some other test subjects reported being more liberal with their responses. However, it does not seem that there is a noticeable difference between the subjects because of this, so this may be less important.

The next section that may benefit from an investigation is the reconstruction and analysis of the responses. The results that we found used linear reconstructions as their basis. However, compressive sensing was also used to see if there was any validity to that reconstruction method. We found that it was not valuable as a reconstruction method, but that comes with a caveat; compressive sensing is based on several variables that were assigned values based on educated guesses. If these variables could be optimized, compressive sensing could prove a valuable method for reconstruction, allowing for reduced storage space and improved processing speed.

The final aspect of the design space that could be altered is the experiment's setup. Due to the nature of this experiment, it again suffered from a lack of standardization. The experiments were done in various environments (none of which were soundproof), and the equipment, such as headphones, varied between proctors. These experiments should be done in a more professional setting, with higher-quality headphones, to get more precise results. However, it may be obvious that asking test subjects to come into a research facility may prove more difficult than finding subjects conveniently (e.g., cohabitants, coworkers, friends).

6.2 Modifications for Clinical Assay

As mentioned previously, the end goal of this project is to create a clinical assay for patients experiencing tinnitus. Since this experiment was done for control purposes, a few modifications would need to be made for this to be appropriate for clinical purposes.

Firstly, and likely most obvious, the control sound would need to be removed. Currently, the test subjects compare the stimulus sound to a control sound. Control sounds are unnecessary

for a clinical setting, as the patient's tinnitus tone would replace the control sound and function as the basis for a patient's perceptual assessment and ultimate response to the stimuli.

Secondly, the analysis of clinical patients would need to be approached differently. The accuracy of the reconstruction would need to be assessed based on model-driven response prediction, which has been used in RC experiments where no control signal is present (cite). Moreover, the analysis methods would need to be augmented to reflect clinical significance rather than statistical significance. For example, the quality of the tinnitus reconstruction could be based on the efficacy of its incorporation into sound therapy or other treatment protocols.

6.3 Miscellaneous Considerations

Finally, some other aspects of this experiment would benefit from amendments in future iterations. Firstly is the demographics of the testing group. As you can see in Section 4, many of the test subjects were white and college-aged. The race of the test subjects may not be of great importance, but experiencing tinnitus is strongly correlated with age. This means that most people who use this assay in a clinical setting will be older. Therefore, it would be useful to use a wider age range of individuals when testing the control.

Another factor to consider is the target sounds that were used as controls. The ATA provides several tones common for patients with tinnitus, and this experiment only tested two of them, roaring and buzzing. Table 7 in Section 4.5 shows that there is nearly no statistical difference between the two sounds that we used; however, this is not necessarily true for all of the sounds that the ATA provides, so for the sake of thoroughness, the next experiment to test this may want to consider increasing the number of test sounds.

7. Conclusion

In conclusion, incorporating Stimulus Generation D into a clinical setting would improve the efficacy of a patient's constant non-tonal PTS characterization. The study's results indicate that incorporating this into clinical settings would improve the treatment of patients with tinnitus due to the reliance and improvement of current therapies on accurate characterizations. While there remain some avenues to be pursued to improve upon some factors of this experiment, such as an increased number of test subjects, alterations to other components of the design space, or standardized testing environments, we believe that the implications of this work provide a solid foundation for clinical characterization of continuous, non-tonal tinnitus.

8. References

Anatomy of the Ear | Inner Ear | Middle Ear | Outer Ear. (n.d.). Retrieved October 18, 2022, from https://www.earq.com/hearing-health/anatomy-of-the-ear

Baguley, D., McFerran, D., & Hall, D. (2013). Tinnitus. *Lancet (London, England)*, *382*(9904), 1600–1607. https://doi.org/10.1016/S0140-6736(13)60142-7

Bhatt, J. M., Lin, H. W., & Bhattacharyya, N. (2016). Prevalence, Severity, Exposures, and Treatment Patterns of Tinnitus in the United States. *JAMA Otolaryngology-- Head & Neck Surgery*, *142*(10), 959–965. https://doi.org/10.1001/jamaoto.2016.1700

Biswas, R., Lugo, A., Genitsaridi, E., Trpchevska, N., Akeroyd, M. A., Cederroth, C. R., Liu, X., Schlee, W., Garavello, W., Gallus, S., & Hall, D. A. (2021). Chapter 1 - Modifiable lifestyle-related risk factors for tinnitus in the general population: An overview of smoking, alcohol, body mass index and caffeine intake. In B. Langguth, T. Kleinjung, D. D. Ridder, W. Schlee, & S. Vanneste (Eds.), *Progress in Brain Research* (Vol. 263, pp. 1–24). Elsevier. https://doi.org/10.1016/bs.pbr.2021.04.005

Brimijoin, W. O., Akeroyd, M. A., Tilbury, E., & Porr, B. (2013). The internal representation of vowel spectra investigated using behavioral response-triggered averaging. *The Journal of the Acoustical Society of America*, *133*(2), EL118–EL122. https://doi.org/10.1121/1.4778264

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. European Review of Social Psychology, 28(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brockmann, L. C. and A. (2005). *Perception Space—The Final Frontier*. https://click.endnote.com/viewer?doi=10.1371%2Fjournal.pbio.0030137&token=WzI4

NDk5MjYsIjEwLjEzNzEvam91cm5hbC5wYmlvLjAwMzAxMzciXQ.mZvG-Tg8eN6cejJjmLp6hhkNAAY

Cederroth, C. R., Gallus, S., Hall, D. A., Kleinjung, T., Langguth, B., Maruotti, A., Meyer, M., Norena, A., Probst, T., Pryss, R., Searchfield, G., Shekhawat, G., Spiliopoulou, M., Vanneste, S., & Schlee, W. (2019). Towards an Understanding of Tinnitus Heterogeneity. *Frontiers in Aging Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnagi.2019.00053

Davis, P. B., Paki, B., & Hanley, P. J. (2007). Neuromonics Tinnitus Treatment: Third clinical trial. *Ear and Hearing*, *28*(2), 242–259. https://doi.org/10.1097/AUD.0b013e3180312619

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Gosselin, F., & Schyns, P. G. (2003). Superstitious Perceptions Reveal Properties of Internal Representations. *Psychological Science*, *14*(5), 505–509. https://doi.org/10.1111/1467-9280.03452

Grewal, R., Spielmann, P. M., Jones, S. E. M., & Hussain, S. S. M. (2014). Clinical efficacy of tinnitus retraining therapy and cognitive behavioural therapy in the treatment of subjective tinnitus: A systematic review. *The Journal of Laryngology and Otology*, *128*(12), 1028–1033. https://doi.org/10.1017/S0022215114002849

Henry, J. A., Flick, C. L., Gilbert, A., Ellingson, R. M., & Fausti, S. A. (2004). Comparison of manual and computer-automated procedures for tinnitus pitch-matching. *Journal of Rehabilitation Research and Development*, *41*(2), 121–138. https://doi.org/10.1682/jrrd.2004.02.0121

Hoyland, A., Barnett, N. V., Roop, B. W., Alexandrou, D., Caplan, M., Mills, J., Parrell, B., Chari, D. A., & Lammert, A. C. (2023). Reverse Correlation Uncovers More Complete Tinnitus Spectra (p. 2022.12.23.521795). bioRxiv. https://doi.org/10.1101/2022.12.23.521795

Jastreboff, P. J. (2011). Tinnitus Retraining Therapy. In *Tinnitus* (pp. 575–596).

Kaylie, D. M. (2022). *Ear Ringing or Buzzing—Ear, Nose, and Throat Disorders*. Merck Manuals. https://www.merckmanuals.com/home/ear,-nose,-and-throat-disorders/symptoms-of-ear-disorders/ear-ringing-or-buzzing

Landgrebe, M., Azevedo, A., Baguley, D., Bauer, C., Cacace, A., Coelho, C., Dornhoffer, J., Figueiredo, R., Flor, H., Hajak, G., van de Heyning, P., Hiller, W., Khedr, E., Kleinjung, T., Koller, M., Lainez, J. M., Londero, A., Martin, W. H., Mennemeier, M., … Langguth, B. (2012). Methodological aspects of clinical trials in tinnitus: A proposal for an international standard. *Journal of Psychosomatic Research*, *73*(2), 112–121. https://doi.org/10.1016/j.jpsychores.2012.05.002

Larem, A., Al Shawabkeh, M. A., & Omer, W. (2021). Tinnitus and Hyperacusis. In A. Al-Qahtani, H. Haidar, & A. Larem (Eds.), *Textbook of Clinical Otolaryngology* (pp. 121–128). Springer International Publishing. https://doi.org/10.1007/978-3-030-54088-3_11

Listen to Sample Tinnitus Sounds. *ATA*. (2022). URL: https://www.ata.org/listen-sample-tinnitus-sounds

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*, 209–226. https://doi.org/10.1016/j.cogsci.2003.11.004

Mehdi, M., Stach, M., Riha, C., Neff, P., Dode, A., Pryss, R., Schlee, W., Reichert, M., & Hauck, F. J. (2020). Smartphone and Mobile Health Apps for Tinnitus: Systematic Identification, Analysis, and Assessment. *JMIR MHealth and UHealth*, *8*(8), e21767. https://doi.org/10.2196/21767

Moon, K., Kim, S., Kim, J., Kim, H., & Ko, Y.-G. (2020). The Mirror of Mind: Visualizing Mental Representations of Self Through Reverse Correlation. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01149

Moore, B. (2013). The Nature of Sound and the Structure and Function of the Auditory System. In *Introduction to the Psychology of Hearing* (6th ed.). Brill.

NIDCD. (2017). *Tinnitus*. https://www.nidcd.nih.gov/health/tinnitus

Nondahl, D. M., Cruickshanks, K. J., Dalton, D. S., Klein, B. E. K., Klein, R., Schubert, C. R., Tweed, T. S., & Wiley, T. L. (2007). The impact of tinnitus on quality of life in older adults. *Journal of the American Academy of Audiology*, *18*(3), 257–266. https://doi.org/10.3766/jaaa.18.3.7

Okamoto, H., Stracke, H., Stoll, W., & Pantev, C. (2010). Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(3), 1207–1210. https://doi.org/10.1073/pnas.0911268107

Parts and Components of Human Ear and Their Functions. (n.d.). MD-Health.Com. Retrieved October 18, 2022, from https://www.md-health.com/Parts-Of-The-Ears.html

Perry. (2021, May 5). *Ear Anatomy: Understanding the Outer, Middle, and Inner Parts of the Ear*. Audiology Research. https://www.audiologyresearch.org/ear-anatomy/

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

Benjamin W. Roop, Benjamin Parrell, & Adam C. Lammert. (2021). A Compressive Sensing Approach for Inferring Cognitive Representations with Reverse Correlation. *BioRxiv*, 2021.09.02.458720. https://doi.org/10.1101/2021.09.02.458720

Sanchez, L. (2004). The epidemiology of tinnitus. *Audiological Medicine*, *2*(1), 8–17. https://doi.org/10.1080/16513860410027781

Schaette, R., König, O., Hornig, D., Gross, M., & Kempter, R. (2010). Acoustic stimulation treatments against tinnitus could be most effective when tinnitus pitch is within the stimulated frequency range. *Hearing Research*, *269*(1–2), 95–101. https://doi.org/10.1016/j.heares.2010.06.022

Shekhawat, G. S., Searchfield, G. D., & Stinear, C. M. (2014). The relationship between tinnitus pitch and hearing sensitivity. *European Archives of Oto-Rhino-Laryngology*, *271*(1), 41–48. https://doi.org/10.1007/s00405-013-2375-6

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology: CB*, *22*(3), 191–196. https://doi.org/10.1016/j.cub.2011.11.061

Tunkel, D. E., Bauer, C. A., Sun, G. H., Rosenfeld, R. M., Chandrasekhar, S. S., Cunningham, E. R., Archer, S. M., Blakley, B. W., Carter, J. M., Granieri, E. C., Henry, J. A., Hollingsworth, D., Khan, F. A., Mitchell, S., Monfared, A., Newman, C. W., Omole, F. S., Phillips, C. D., Robinson, S. K., … Whamond, E. J. (2014). Clinical Practice Guideline: Tinnitus. *Otolaryngology–Head and Neck Surgery*, *151*(2), S1–S40. https://doi.org/10.1177/0194599814545325

Umesh, S., Cohen, L., & Nelson, D. (1999). *Fitting the Mel scale* (Vol. 1, p. 220 vol.1). https://doi.org/10.1109/ICASSP.1999.758101

Wang, H., Tang, D., Wu, Y., Zhou, L., & Sun, S. (2020). The state of the art of sound therapy for subjective tinnitus in adults. *Therapeutic Advances in Chronic Disease*, *11*, 2040622320956426. https://doi.org/10.1177/2040622320956426

Yew, K. S. (2014). Diagnostic Approach to Patients with Tinnitus. *American Family Physician*, *89*(2), 106–113.

Zenner, H.-P., Vonthein, R., Zenner, B., Leuchtweis, R., Plontke, S. K., Torka, W., Pogge, S., & Birbaumer, N. (2013). Standardized tinnitus-specific individual cognitive-behavioral therapy: A controlled outcome study with 286 tinnitus patients. *Hearing Research*, *298*, 117–125. https://doi.org/10.1016/j.heares.2012.11.013

9. Appendix

9.1 Config File

```
# This is a template config file that describes how to customize it.
# It is a fully-functional config file on its own,
# but includes additional comments to help explain what the different
# fields mean.

# A good syntax for this field is
# {experiment_name}-{target_signal}-{subject_initials}-[resynth]
experiment_name: StimGenA-buzzing-MC
subject_ID: MC

# These fields describe the number of trials in the experiment.

# n_trial_per_block is the number of trials per block of the experiment.
# A block is a set of contiguous trials without a break.
# Subjects get a break between blocks.
# These are both required fields.
n_trials_per_block: 100
n_blocks: 20

# The total trials should be the number of trials per block
# times the number of blocks.
# This is not a required field.
total_trials: 2000

# These "freq" fields describe the frequency range of the stimuli,
# including the minimum frequency and maximum frequency,
# both in Hz.
# The duration field describes the duration
# of the stimulus in seconds.
# These are not required fields.
# Default values are set to min_freq = 100 and max_freq = 22000
# and duration = 0.5.
# These defaults are defined in
# tinnitus-project/code/stimulus-generation/@AbstractStimulusGenerationMethod.
min_freq: 100
max_freq: 13000
duration: 0.5

# For a stimulus type that uses bins,
# the number of bins are set here.
# This should be a positive scalar integer.
n_bins: 100

# This required parameter gives the stimuli type.
# The name is the class that defines the stimuli type
```

```
# without "StimulusGeneration".
stimuli_type: UniformPrior

# Some stimulus generation methods have other parameters
# associated with them.
# For example, the Gaussian Prior stimulus generation method
# requires an n_bins_filled_mean and n_bins_filled_var property.
# You can see what extra parameters are required for your method
# by inspecting the class definition for the method,
# e.g., at tinnitus-project/code/stimulus-generation/.
# If you do not overwrite values in the config,
# default values are used, which are described
# in the class definition.
min_bins: 30
max_bins: 30

# For an experiment with a target signal
# (i.e., for pilot subjects)
# this field describes the full filepath
# to the target signal audio file.
target_signal_filepath:
C:\Users\myahc\Documents\GitHub\tinnitus-project\code\experiment\ATA\ATA_Tinnit
us_Buzzing_Tone_1sec.wav

# This field gives the short-form name of the target signal.
# For resynth experiments, you should name it
# "resynth-buzzing" for instance.
target_signal_name: buzzing

# This field is a boolean flag
# that indicates whether the target signal
# should be binned (and then unbinned)
# before playback to the user.
bin_target_signal: true

# This is the path where the output files are saved.
# This is not a required field.
# If it is is unset, it will default to
# tinnitus-project/code/experiment/Data.
data_dir: C:\Users\myahc\Documents\GitHub\StimGenAWeek1\MCDataNov1

# This field determines in what form the stimuli are saved.
# The available options are 'bins', 'waveform', or 'spectrum'.
# If not set, it will default to 'waveform'.
stimuli_save_type: bins
```

9.2 Consent Form

UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL

COMMITTEE FOR THE PROTECTION OF HUMAN SUBJECTS IN RESEARCH

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Title: H00024058 - Tinnitus

Investigator: Divya Chari

908 Windsor Ridge Dr.

Westborough, MA, 01581

Daytime Phone Number: 612-747-6308

Consent Version: August 25th, 2021

You are being invited to take part in a research study. Someone will explain this research to you. This form helps to sum up their explanation.

KEY INFORMATION

You are being invited to participate in a research study because you experience tinnitus. Tinnitus is when you experience ringing or other noises in one or both ears. The noise you hear when you have tinnitus is not caused by an external sound and other people usually cannot hear it.

If you have questions or don't understand something, please ask.

Taking part in this research is voluntary and completely up to you. You are free to say no or to leave the research at any time. There will be no penalties or changes in the quality of the health care you receive, and you will not lose any benefits to which you are otherwise entitled.

The main question this study is trying to answer is whether the tinnitus that individuals experience can be recreated from a series of random sounds.

If you join this research, you may undergo a comprehensive audiological examination and be asked to listen to a series of randomly generated excerpts of white noise over a period of 2-10 hours. You may also be asked to complete one or more surveys. In addition, we may ask you information about your otologic (ear) medical history or collect information about your otologic history from your medical record.

You may not want to be in this study if you are uncomfortable with:

- Talking about or explaining your tinnitus to others
- Sharing your private information with researchers

Risks: There are minimal risks to participating in this study. There is a small risk of breach of confidentiality and a risk of fatigue and the possibility that we will need to schedule tasks over multiple sessions. We will take steps to protect your personal information. You will be assigned a de-identified subject number and all information that is collected will be stored under a de-identified database.

Benefits: Although there are no direct benefits to you for participating in this study, your participation may help us to gain knowledge that to help patients with tinnitus in the future.

Alternatives: You do not have to be in this study to receive medical care for your tinnitus or hearing loss. Participating in this study will not change the care that you receive at any medical institution.

Conflict of Interest: There are no conflicts of interest.

If you think you might like to participate in this research, please continue reading to learn more about the details of this study.

STUDY DETAILS

How many people will take part in this research?

About 100 people will take part in this study at UMass Memorial Center and Worcester Polytechnic Institute.

How long will I be in this research?

The study will take about 2-10 hours to complete. No session will be longer than 2 hours, but we may ask you to return for additional sessions. There will be frequent breaks.

What happens if I say yes, I want to be in this research?

If you decide to participate in this research, you may be asked to complete surveys or questionnaires to describe the severity of your hearing loss and/ or tinnitus.

You will be asked to listen to a series of excerpts of white noise and try to match the white noise to your perception of your tinnitus.

The study will take place at Worcester Polytechnic Institute.

Will you be collecting any specimens from me?

No.

Will being in this research help me in any way?

Unfortunately, we do not expect that participating in this study will improve or change your tinnitus. However, participating in this study may allow researchers to better understand tinnitus so that we may one day help improve treatments for tinnitus.

What other choices do I have besides taking part in this research?

You do not have to participate in this research. If, at any point, you chose to stop participating in this research, you may do so without any adverse consequences.

Will it cost me any money to take part in this research?

No, there is no cost to you.

Will I be given any money or other compensation for being in this study?

You will be paid $10/ hour for participation in the study and you will be compensated for parking.

What happens if I am injured because I took part in this research?

We do not expect any injuries from participation in this study. However, if you are injured while in the study, seek treatment and contact the study doctor as soon as you are able.

The University of Massachusetts Medical School and Worcester Polytechnic Institute do not provide funds for the treatment of research-related injury. If you are injured as a result of your participation in this study, treatment will be provided. You or your insurance carrier will be

expected to pay the costs of this treatment. No additional financial compensation for injury or lost wages is available.

You do not give up any of your legal rights by signing this form.

What happens if I say yes, but I change my mind later?

If you decide not to participate in this research study at any point during testing, please let us know. There will be no adverse consequences to you for withdrawing from this study.

How will my information be stored and when will it be destroyed?

We will remove your name and any other information that could directly identify you from your data. We will replace this information with a code number. We will create a master list linking your code number to your name. We will keep this list separate from your data.

We will keep paper documents under lock and key. We will keep electronic health information and research data on secure computer networks. These computer networks have many levels of protection.

There is no limit on the length of time we will store your data.

- It is possible that we might use the research data and specimens in other future research. We may also share data and specimens with researchers and companies that are not part of UMMS. In these cases, we will not share your name or other information that identifies you directly, and we will not come back to you to ask you for your consent.

Who has access to my information?

Signing this document means you allow us, the researchers in this study, and others working with us to use some protected health information for this research study.

As part of the research, UMass Memorial Medical Center or any other healthcare facility where you are treated may disclose the following information:

- Demographic and identifying information like your name, date of birth, address, telephone number, and your email address
- Related medical information like family medical history, hearing-related information, and current and past medications or therapies
- All tests and procedures that will be done in the study

Your health information and research records will be shared with the study team and with individuals and organizations that conduct or watch over this research, in order to conduct the study and to make sure it is conducted as described in this form. Information and records may be shared with:

- The Institutional Review Board (IRB) that reviewed this research
- The University of Massachusetts Medical School and UMass Memorial Health Care, including their Institutional Review Board (IRB) and research, billing, and compliance offices
- Health care providers who provide services in connection with this study

We will protect your identifiable information from disclosure to others to the extent required by law, but we cannot promise complete secrecy.

Will you share any results with me?

We will share the results of the audiologic examination with you. The other tests that are conducted at WPI can be shared with you if you ask. They cannot tell you about your health or diagnose any condition.

Who can I talk to?

If you have questions, concerns, or complaints, or think this research has hurt you or made you sick, talk to the research team at the phone number listed on the first page.

This research is being overseen by an Institutional Review Board. An IRB is a group of people who perform independent review of research studies. You may talk to them at (508) 856-4261 or irb@umassmed.edu for any of the following:

Your questions, concerns, or complaints are not being answered by the research team.

You cannot reach the research team.

You want to talk to someone besides the research team.

You have questions about your rights as a research participant.

You want to get information or provide input about this research.

Signature Block for Capable Adults

Your signature documents your consent to take part in this research.

_____ _____

Signature of adult research participant Date

_____

Printed name of adult research participant

_____ _____

Signature of person obtaining consent Date

_____

Printed name of person obtaining consent

9.3 Table of R and R^2 Values of Each Test Subject

Stim Gen	Sound	ID	Time (min)	R Values		R^2 Values	
				LR	CS	LR	CS
A UniformPrior	Buzzing	Subject 1	45	0.341	0.576	0.116	0.331
		Subject 2	27	0.275	0.566	0.075	0.321
		Subject 3	26	-0.230	-0.370	0.053	0.137
		Subject 4	30	0.116	0.259	0.013	0.067
	Roaring	Subject 5	60	0.292	0.366	0.085	0.134
		Subject 6	25	0.089	0.190	0.008	0.036
		Subject 7	30	0.028	0.020	0.001	0.000
		Subject 8	25	0.100	0.032	0.010	0.001
		Mean	33.5	0.126	0.205	0.045	0.128
		STDev	12.5	0.183	0.314	0.043	0.133
		CoV %	37.5	145.0	153.4	96.1	103.2
B UniformPrior8	Buzzing	Subject 1	30	0.755	0.831	0.570	0.691
		Subject 9	30	-0.018	0.106	0.000	0.011
		Subject 4	30	0.361	0.831	0.130	0.691
		Subject 10	32	0.878	0.831	0.770	0.691
	Roaring	Subject 5	30	0.239	0.457	0.057	0.209
		Subject 11	40	0.734	0.457	0.539	0.209
		Subject 12	30	0.158	0.490	0.025	0.240
		Subject 13	18	0.903	0.247	0.815	0.061
		Mean	30.0	0.501	0.531	0.363	0.350
		STDev	6.0	0.358	0.279	0.346	0.293
		CoV %	19.8	71.5	52.5	95.2	83.5
C Brimijoin8	Buzzing	Subject 1	26	0.766	0.831	0.587	0.691
		Subject 2	30	0.825	0.831	0.680	0.691
		Subject 14	34	0.867	0.831	0.751	0.691
		Subject 10	28	0.131	-0.204	0.017	0.042
	Roaring	Subject 15	17	0.158	0.247	0.025	0.061
		Subject 5	21	0.000	0.247	0.000	0.061
		Subject 16	25	0.646	0.632	0.417	0.399
		Subject 17	17	0.552	0.632	0.305	0.399
		Mean	24.8	0.493	0.506	0.348	0.380

		STDev	6.1	0.346	0.376	0.310	0.295
		CoV %	24.6	70.2	74.2	89.1	77.8
D Brimijoin Gaussian Smoothed	Buzzing	Subject 1	25	0.838	---	0.702	---
		Subject 18	30	0.378	---	0.143	---
		Subject 2	30	0.643	---	0.413	---
		Subject 10	36	0.264	---	0.070	---
	Roaring	Subject 5	30	0.674	---	0.454	---
		Subject 6	22	0.883	---	0.780	---
		Subject 19	32	0.836	---	0.699	---
		Subject 20	24	0.392	---	0.154	---
		Mean	28.6	0.614	---	0.427	---
		STDev	4.6	0.240	---	0.282	---
		CoV %	16.2	39.1	---	66.2	---

9.4 Gaussian Target Signal Binned Approximation

```
%% Import Data

T = readtable('GaussianTableT');

reconlr = T(:,23:30);

%% Gaussian Spectrum

minfreq = 100;
maxfreq = 13000;
nfft = maxfreq;
freq = linspace(minfreq,maxfreq,nfft)';
nbins_freq = 8;

% Gaussian Parameters
MU = [150 550 1100 1900 3100 4800 7300 10900]';
SIGMA = [150 250 350 500 700 1000 1500 2150]';

% Example Spectrum
wf = audioread('ATA_Tinnitus_Buzzing_Tone_1sec.wav');
pxx = 10*log10(pwelch(wf));
y = resample(pxx(1:5000),13000,5000);
y = rescale(y);
y = medfilt1(y,300);

% Populate X with Gaussian Basis Functions
X = zeros(nfft,nbins_freq);
for itor = 1:nbins_freq
    X(:,itor) = normpdf(freq,MU(itor),SIGMA(itor));
end

% Fit Gaussians
% These values would be the ones to correlated against when evaluating the
% quality of a reconstruction. These represent the heights of Gaussians
% that best represent the spectrum contained in 'y'.
b_hat = (X'*X)\(X'*y);

% Viz Gaussian Heights
figure
stem(b_hat,'linewidth',2)
xlabel('Gaussian Number')
ylabel('Magnitude')

% Test Gaussian Fit
y_hat = X*b_hat;

% Viz Results
```

```
figure
plot(freq,y,'k','linewidth',2)
hold all
plot(freq,y_hat,'r:','linewidth',2)
xlabel('Frequency')
ylabel('Power')
legend('ATA Spectrum','Gaussian Approx')

%% Compare Target and Subject Data

target = b_hat;
subject = table2array(reconlr)';

Rvals = corr(subject,target);
```

9.5 Statistical Significance Analysis Code

```
%% Import Data

T = readtable('ALLdata');

% Column 1: Stimulus Generation A, B, C, D
Gen = T(:,1);
% Column 2: Target Signal B, R
Targ = T(:,2);
% Column 3: Subject ID
ID = T(:,3);
% Column 4: Time
ID = table2array(T(:,4));
% Column 5: Linear R val
lr = table2array(T(:,5));
% Column 6: Compressive Sensing R val
cs = table2array(T(:,6));

%% ANOVA
% ANOVA 1 assumes 'One-way analysis of variance'
% 1 independent variable, to us it is Stim Gen

% consider fisher transformation before this

%% BUZZING LR

buzlr = [lr(1:4) lr(9:12) lr(17:20) lr(25:28)];
buzlr = atanh(buzlr);

[p1,tbl1,stats1] = anova1(buzlr);
title('ANOVA lr buzzing')

figure
op1 = multcompare(stats1);
title('ANOVA lr buzzing')

% None have any statistically significant difference

%% ROARING LR

roalr = [lr(5:8) lr(13:16) lr(21:24) lr(29:32)];
roalr = atanh(roalr);

[p2,tbl2,stats2] = anova1(roalr);
title('ANOVA lr roaring')

figure
op2 = multcompare(stats2);
```

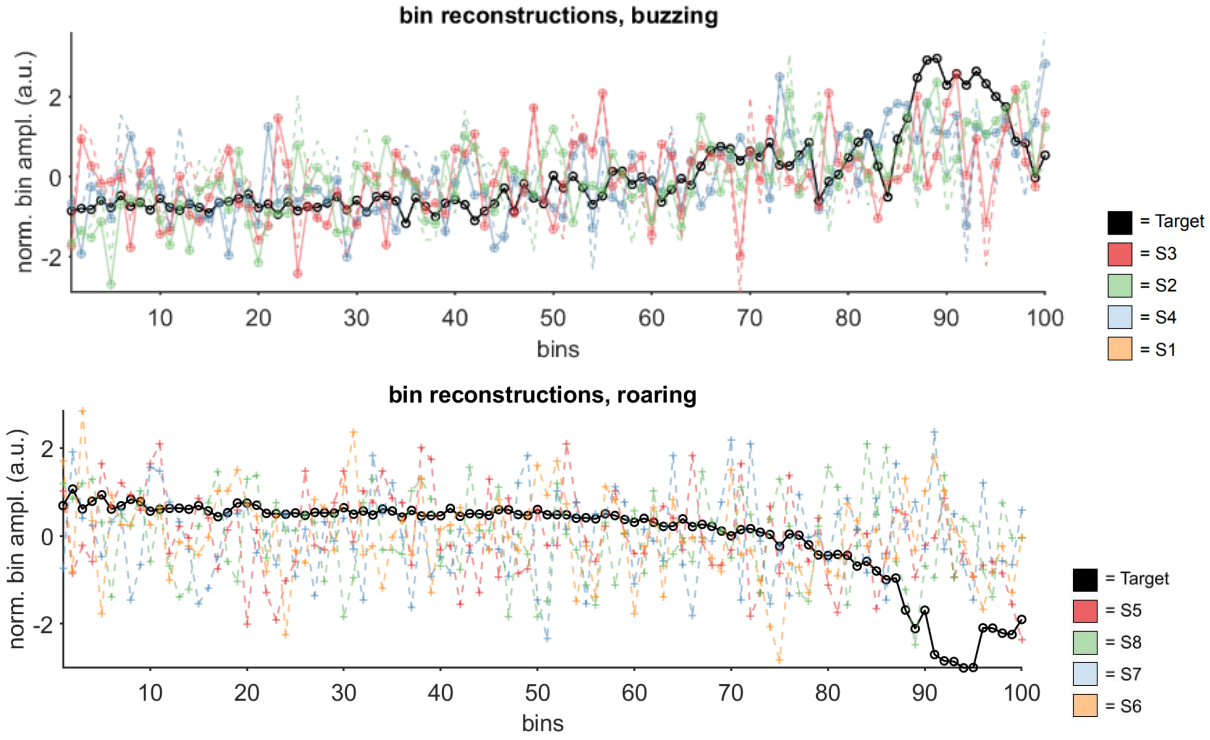
```
title('ANOVA lr roaring')

% The means of group 1 and 4 are statistically different

%% All buzzing vs roaring LR

BvsR = [lr(1:4) lr(5:8); lr(9:12) lr(13:16); lr(17:20) lr(21:24); lr(25:28)
lr(29:32)];
BvsR = atanh(BvsR);

[p3,tbl3,stats3] = anova1(BvsR);
title('ANOVA lr buz vs roa')

figure
op3 = multcompare(stats3);
title('ANOVA lr buz vs roa')

% None have any statistically significant difference

%% All Stim Gen Buz plus Roa LR

banr = [buzlr; roalr];
banr = atanh(banr);

[p4,tbl4,stats4] = anova1(banr);
title('ANOVA All Gen lr Buz an Roa')

figure
op4 = multcompare(stats4);
title('ANOVA All Gen lr Buz an Roa')

% The means of group 1 and 4 are statistically different

%% All Stim Gen Buz vs Roa LR 8

banr8 = [buzlr roalr];
banr8 = atanh(banr8);

[p44,tbl44,stats44] = anova1(banr8);
title('ANOVA All 8 Gen lr Buz vs Roa')

figure
op44 = multcompare(stats44);
title('ANOVA All 8 Gen lr Buz vs Roa')

% None have any statistically significant difference

%% CS
```

```
% BUZZING

buzcs = [cs(1:4) cs(9:12) cs(17:20)];
buzcs = atanh(buzcs);

[p5,tbl5,stats5] = anova1(buzcs);
title('ANOVA cs buzzing')

figure
op5 = multcompare(stats5);
title('ANOVA cs buzzing')

% None have any statistically significant difference

%% ROARING cs

roacs = [cs(5:8) cs(13:16) cs(21:24)];
roacs = atanh(roacs);

[p6,tbl6,stats6] = anova1(roacs);
title('ANOVA cs roaring')

figure
op6 = multcompare(stats6);
title('ANOVA cs roaring')

% None have any statistically significant difference

%% All buzzing vs roaring cs

BvsRcs = [cs(1:4) cs(5:8); cs(9:12) cs(13:16); cs(17:20) cs(21:24)];
BvsRcs = atanh(BvsRcs);

[p7,tbl7,stats7] = anova1(BvsRcs);
title('ANOVA cs buz vs roa')

figure
op7 = multcompare(stats7);
title('ANOVA cs buz vs roa')

% None have any statistically significant difference

%% All Stim Gen Buz plus Roa cs

banrcs = [buzcs; roacs];
banrcs = atanh(banrcs);

[p8,tbl8,stats8] = anova1(banrcs);
title('ANOVA All Gen cs Buz an Roa')
```

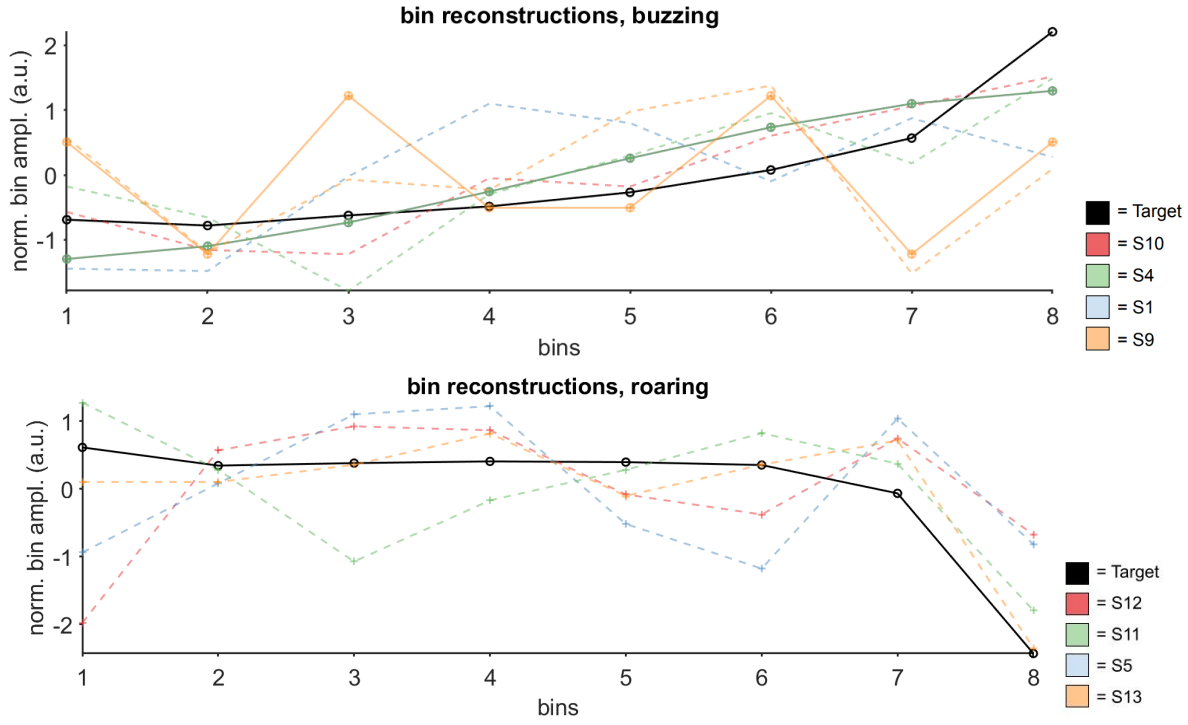
```
figure
op8 = multcompare(stats8);
title('ANOVA All Gen cs Buz an Roa')

% None have any statistically significant difference

%% All Stim Gen Buz vs Roa cs 8

banrcs8 = [buzcs roacs];
banrcs8 = atanh(banrcs8);

[p88,tbl88,stats88] = anova1(banrcs8);
title('ANOVA All 8 Gen cs Buz vs Roa')

figure
op88 = multcompare(stats88);
title('ANOVA All 8 Gen cs Buz vs Roa')

% None have any statistically significant difference
```

9.6 Test Subject Reconstruction Plots

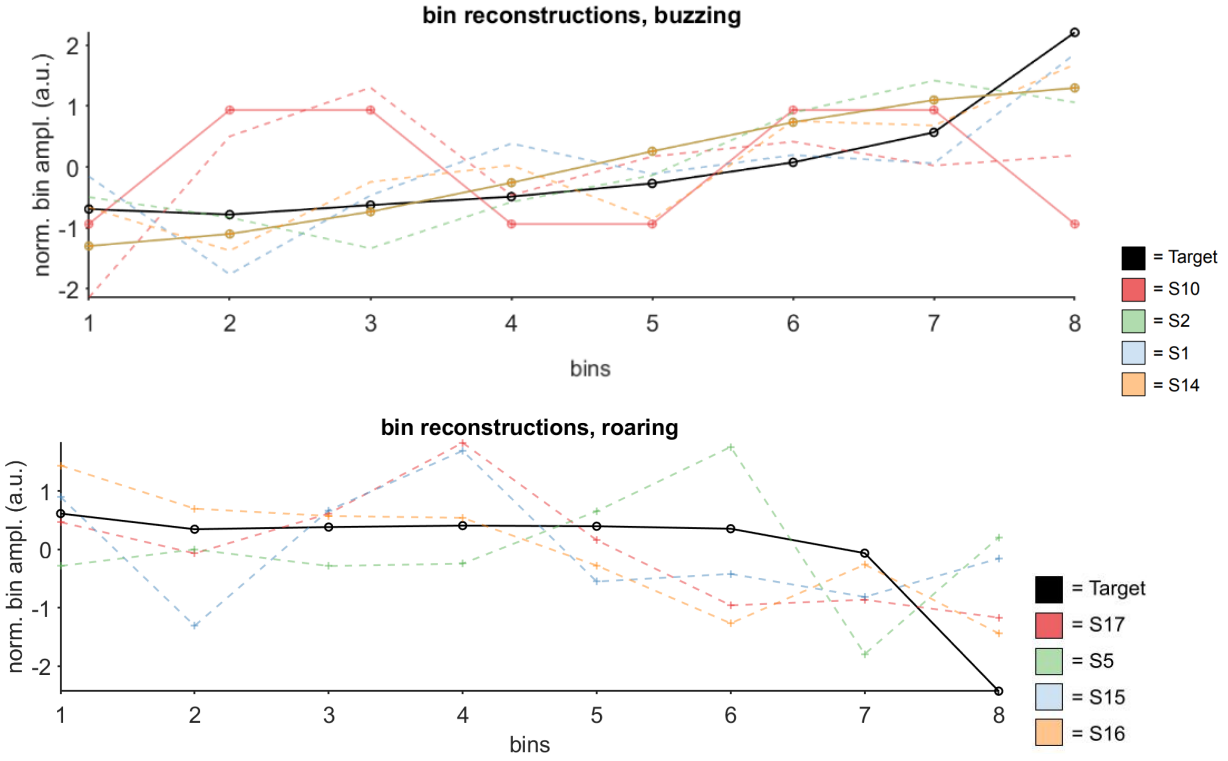
9.6.1 Stimulus Generation A: Uniform Prior



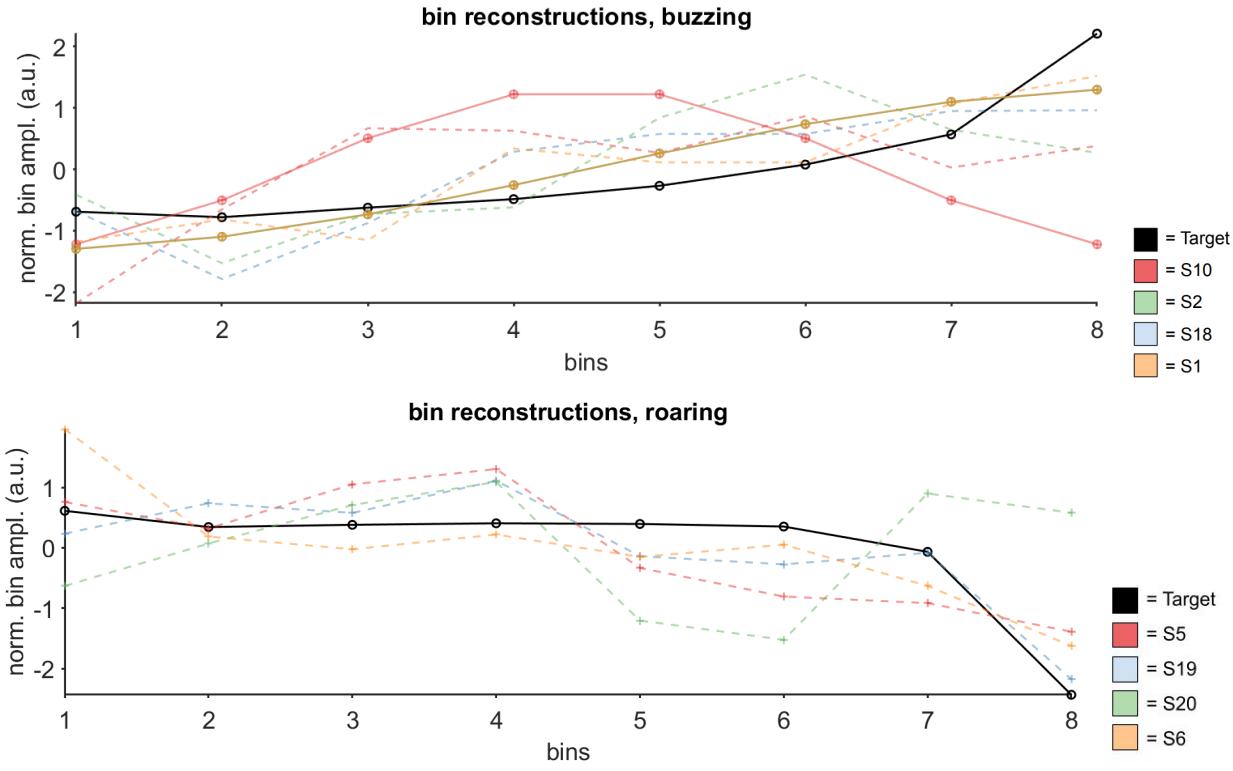
9.6.2 Stimulus Generation B: Uniform Prior 8 Bin



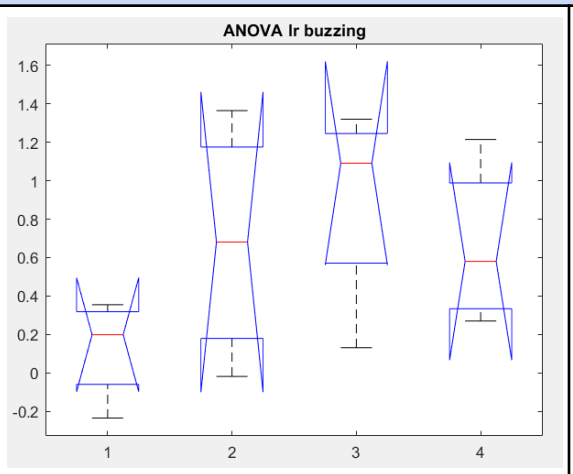
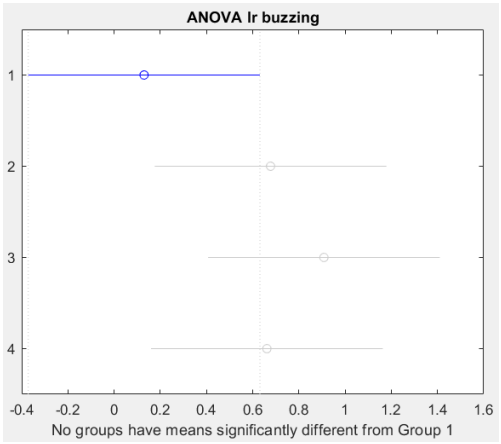
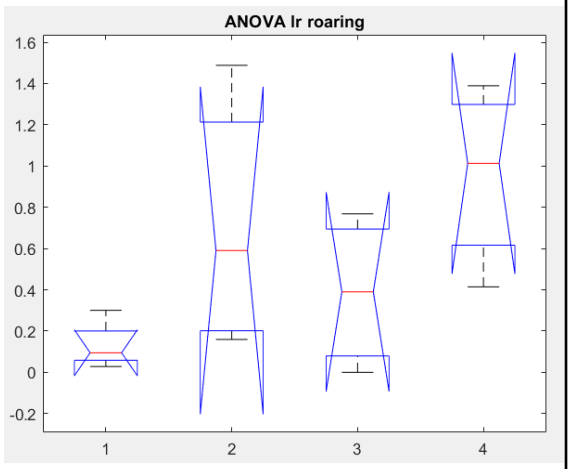
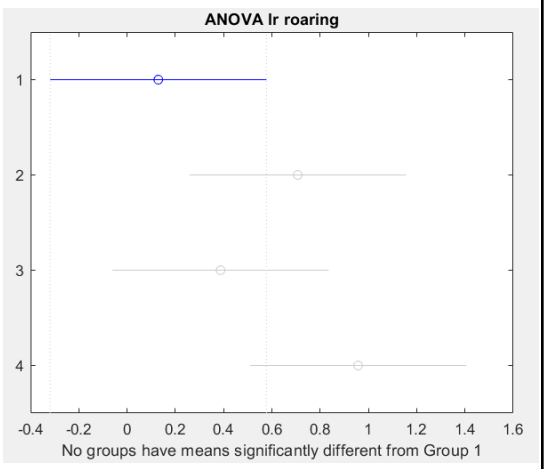
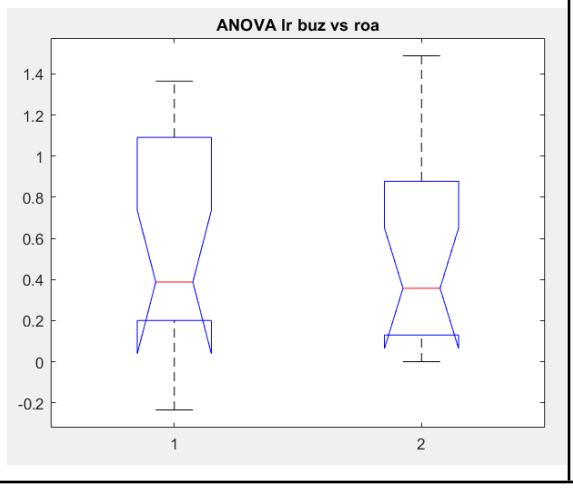
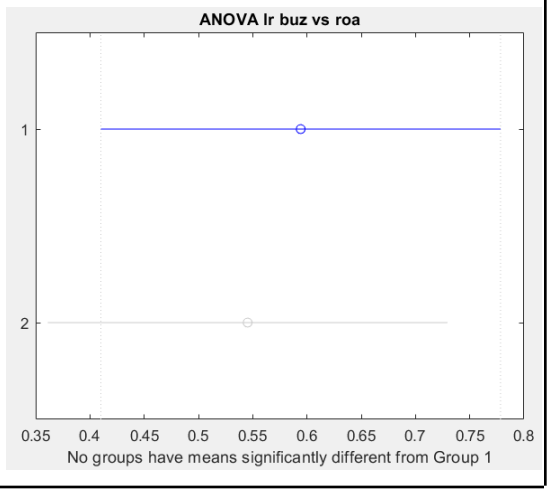
9.6.3 Stimulus Generation C: Brimijoin

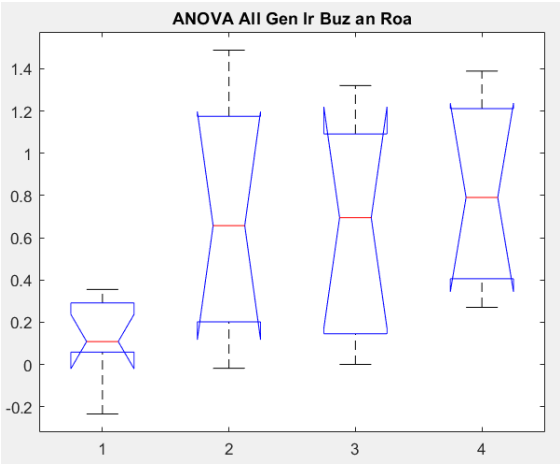
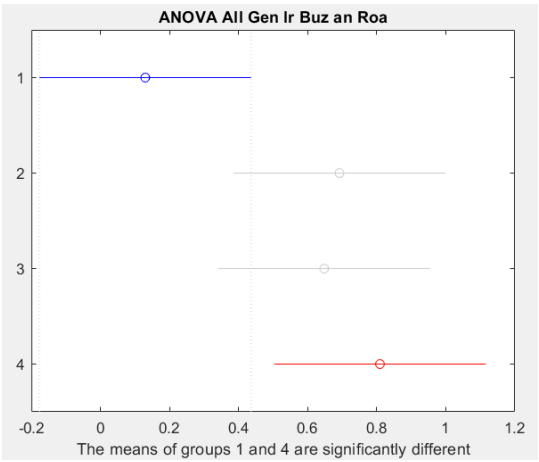
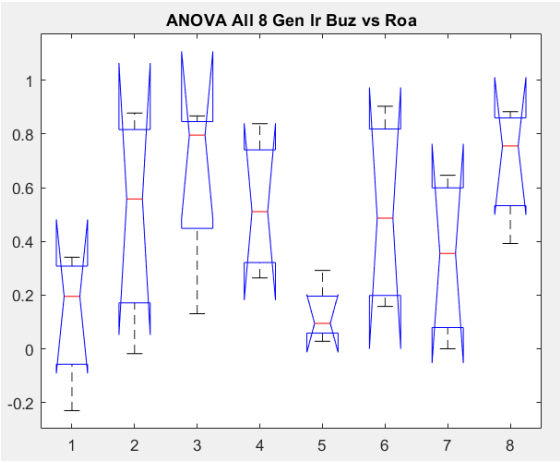
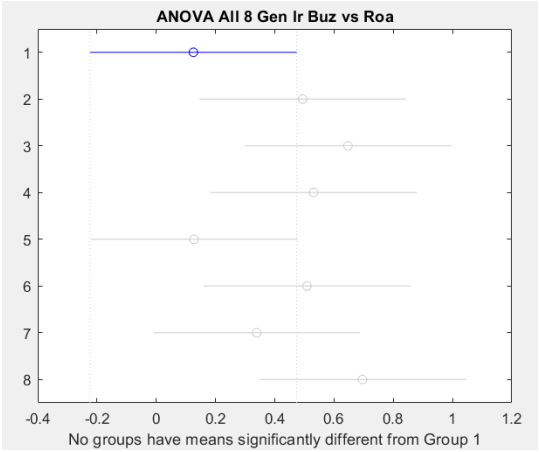
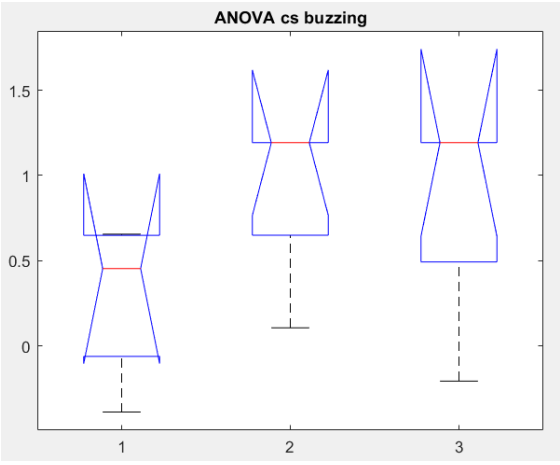
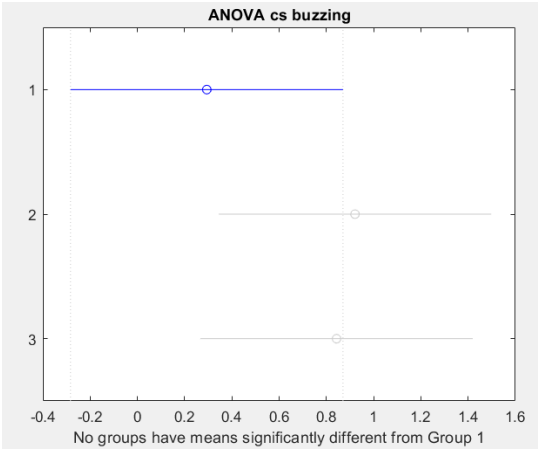


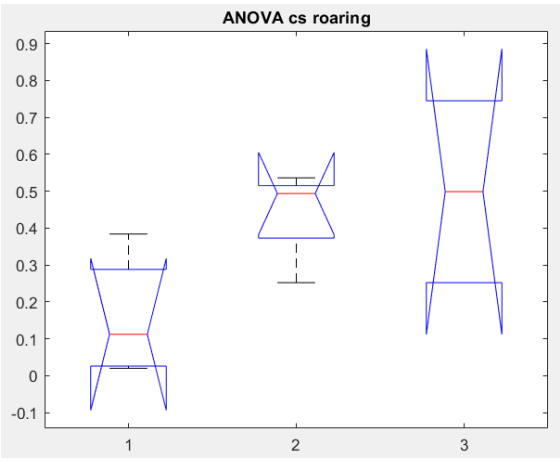
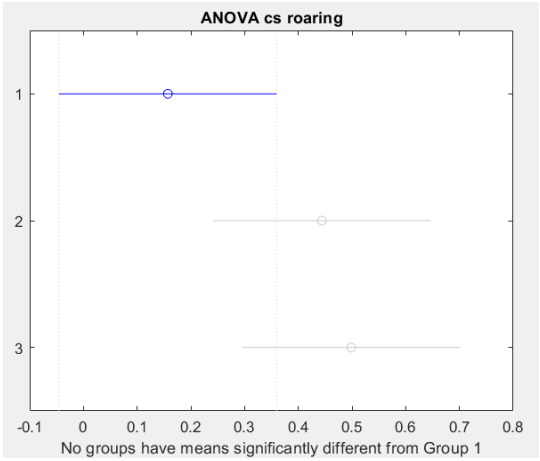
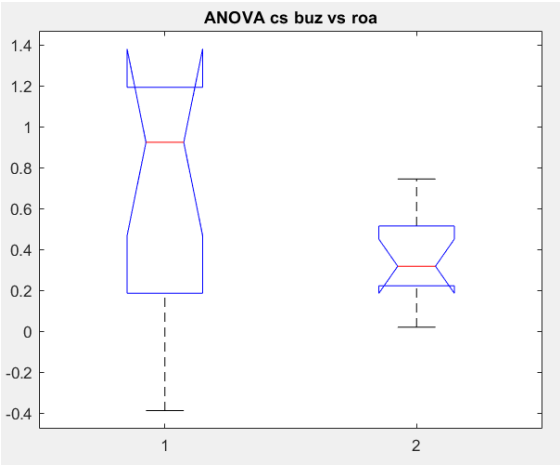
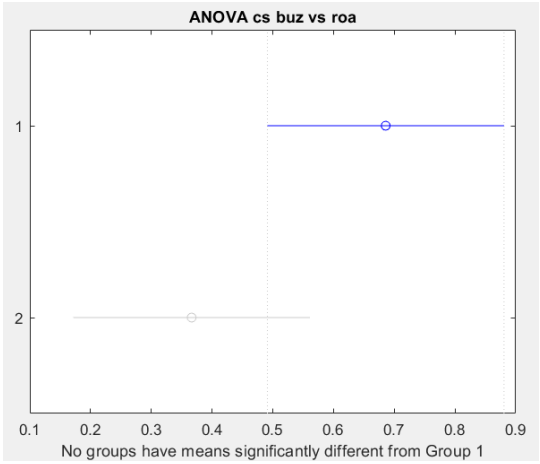
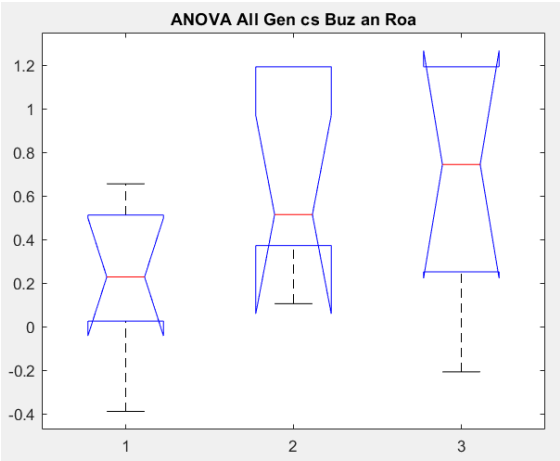
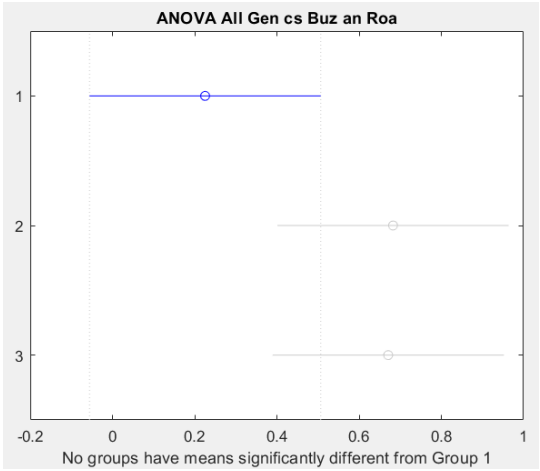
9.6.4 Stimulus Generation D: Brimijoin Gaussian Smoothed

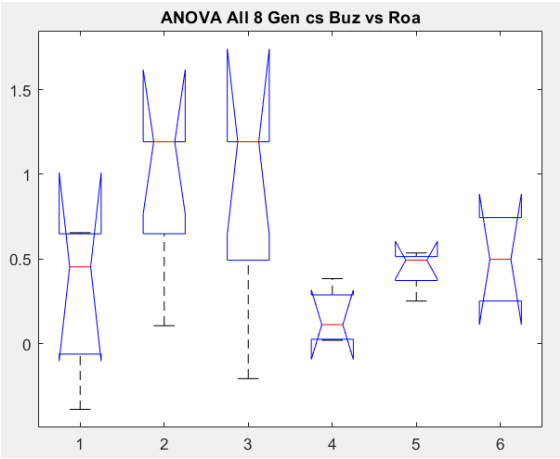
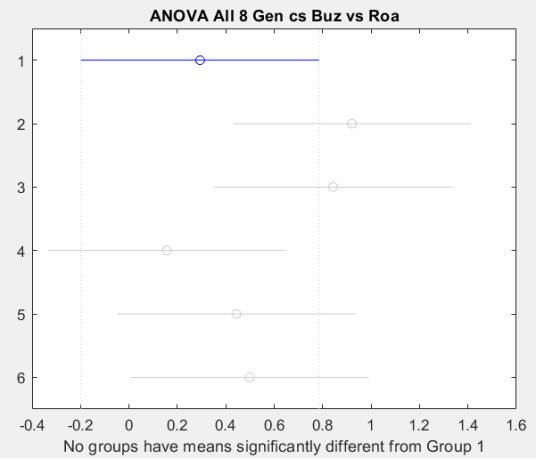


9.7 ANOVA Analysis Results

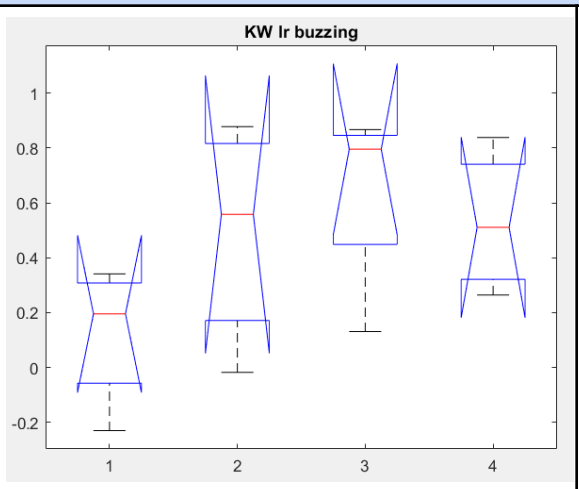
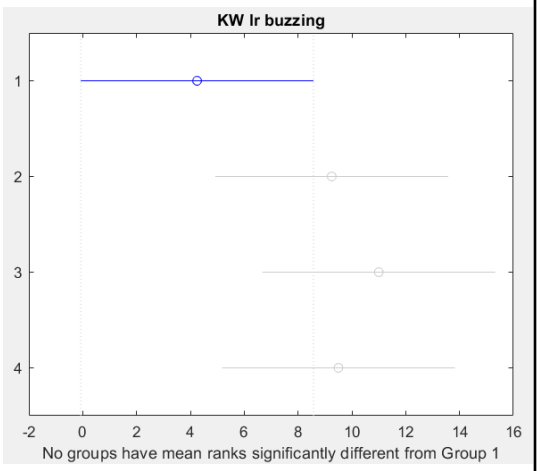
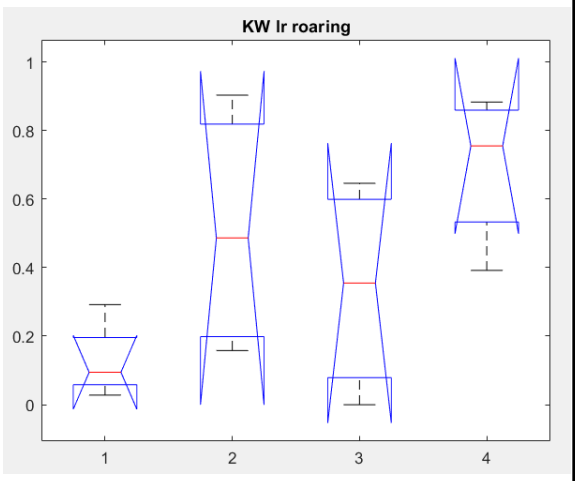
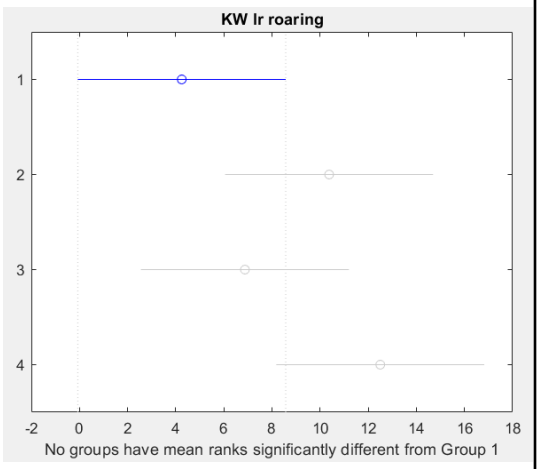
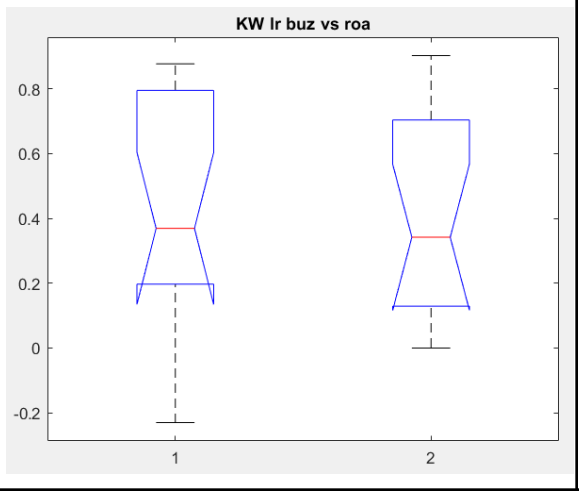
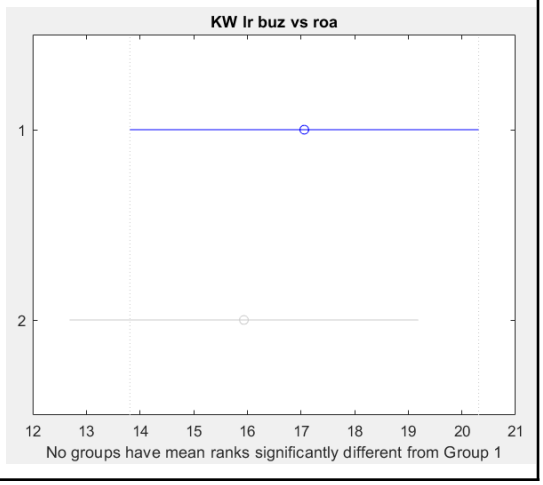
<table>
<tr><th colspan="3">ANOVA LR</th></tr>
<tr><td>A-D only buzzing (4)</td><td> 
</td><td> 
</td></tr>
<tr><td>A-D only roaring (4)</td><td> 
</td><td> 
</td></tr>
<tr><td>Buzzing vs. roaring (2)</td><td> 
</td><td> 
</td></tr>
</table>

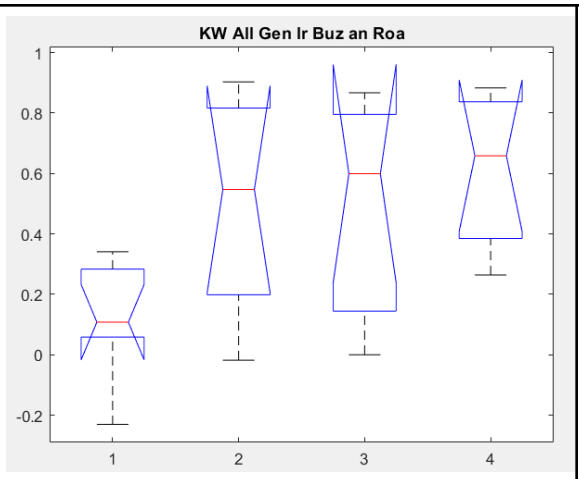
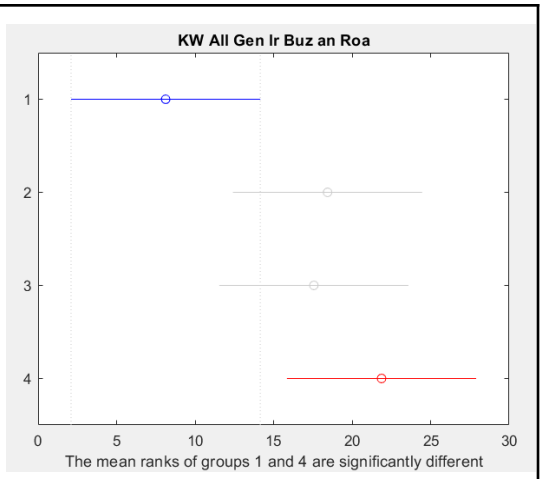
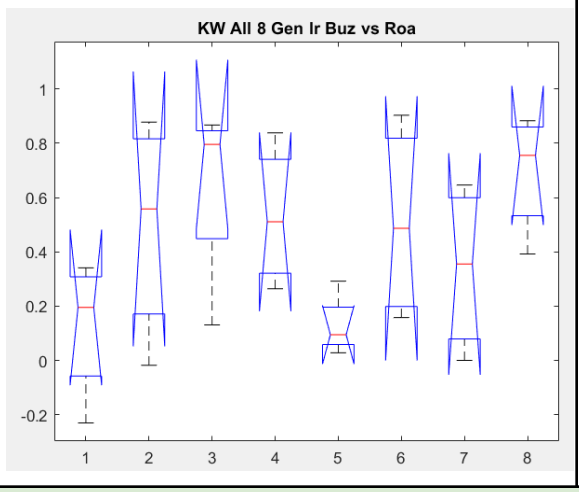
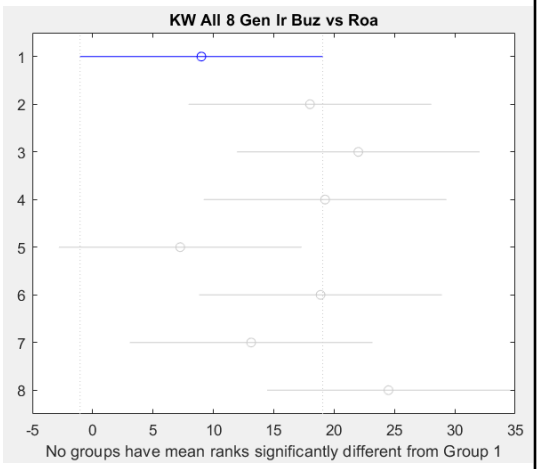
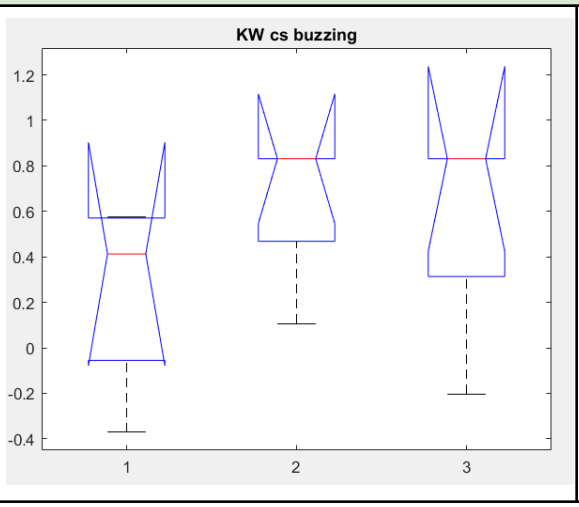
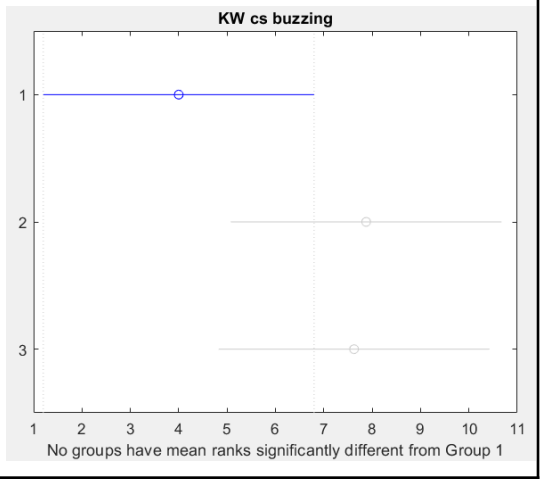
A-D (4)		
A-D buzz vs. roar (8)		
ANOVA CS		
A-D only buzzing (3)		

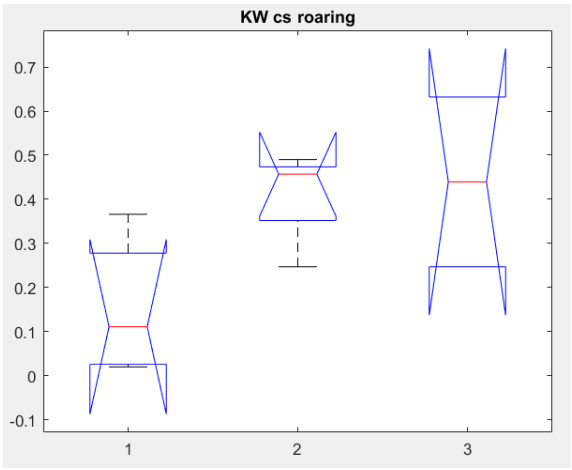
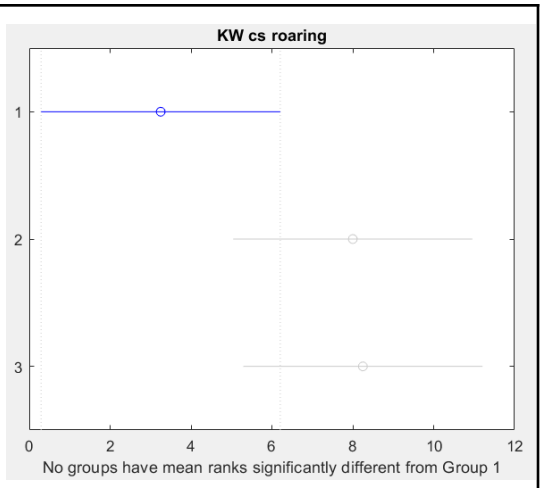
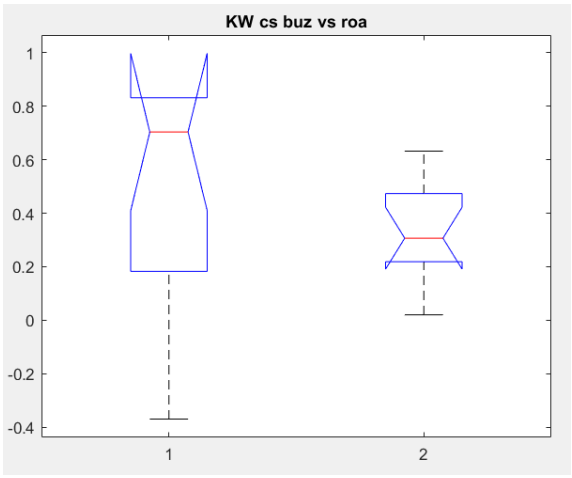
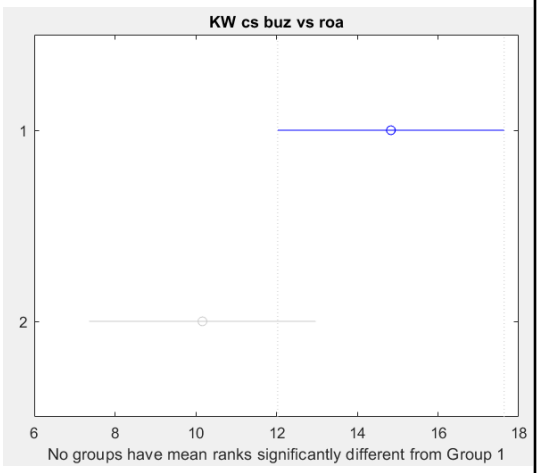
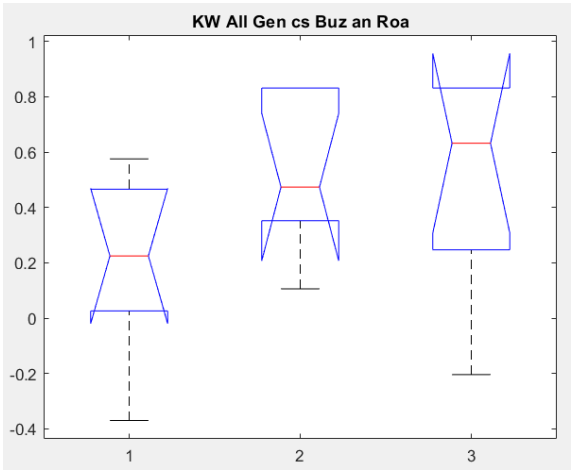
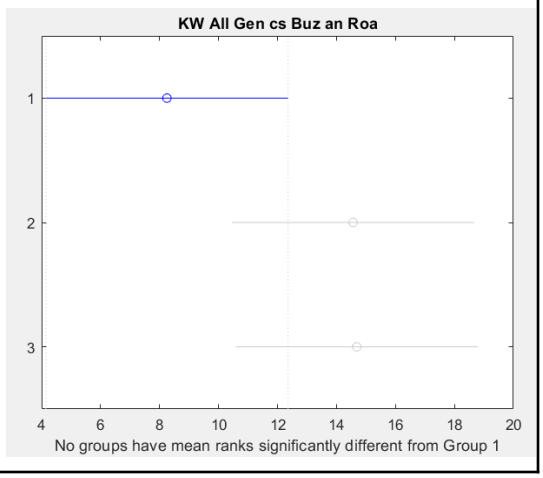
A-D only roaring (3)		
Buzzing vs. roaring (2)		
A-D (3)		

A-D buzz vs. roar (6)		

9.8 Kruskal-Wallis Test

Kruskal Wallis LR		
A-D only buzzing (4)		
A-D only roaring (4)		
Buzzing vs. roaring (2)		

A-D (4)		
A-D buzz vs. roar (8)		
Kruskal Wallis CS		
A-D only buzzing (3)		

A-D only roaring (3)		
Buzzing vs. roaring (2)		
A-D (3)		

A-D buzz vs. roar (6)