USING COMPUTATIONAL OPTIMIZATION FOR AUDIO FILTER DESIGN TO IMPROVE FREQUENCY SELECTIVITY IN THE PASSIVE COCHLEA

A Thesis

by

AARON HOFFMAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Brian Applegate
Committee Members,	Xiaoning Qian
	Vladislav Yakovlev
Head of Department,	Mike McShane

December 2019

Major Subject: Biomedical Engineering

ABSTRACT

Hearing loss affects a significant portion of the global population. It is widely accepted that the effective management of hearing loss early in life helps to avoid development delay, and treating hearing loss later on has a significantly positive effect on quality of life. Conventional hearing aids function by exclusively amplifying sound to compensate for a patient's decreased hearing threshold. However, they do not compensate for the diminished frequency bandwidth that comes along with sensioneural hearing loss. A collection of hydrodynamically-coupled, damped, driven harmonic oscillators are used to simulate the basilar membrane displacement response to an audio signal as a surrogate for directly modeling audio perception. This model, coupled with a high-dimensional, global, gradient-free optimization technique is used to design filtered audio signals which improve the frequency bandwidth in the hearing-damaged cochlea. By substituting an original simple sinusoidal tone with a complex of frequencies that interfere on the cochlea, we show that it is possible to improve frequency selectivity in the damaged cochlea using only an audio filtering technique. This type of process may, in the future, be incorporated with conventional hearing aid technology to further improve the audibility and intelligibility of audio for those with mild hearing loss.

DEDICATION

To my parents.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Brian Applegate for his support and guidance during my time at Texas A&M. I would also like to thank my committee members, Xiaoning Qian and Vladislav Yakovlev for their willingness to help and answer questions that I had during the course of this work.

Finally, I want to thank my colleagues at the Air Force Research Laboratory for giving me the opportunity to attend Texas A&M and supporting me through my time away from the office.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis (or) dissertation committee consisting of Professor Brian Applegate [advisor] and Professor Vladislav Yakovlev of the Department of [Biomedical Engineering] and Professor Xiaoning Qian of the Department of [Electrical & Computer Engineering].

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was supported by the United States Air Force Civil Service Science and Engineering PALACE Acquire program.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

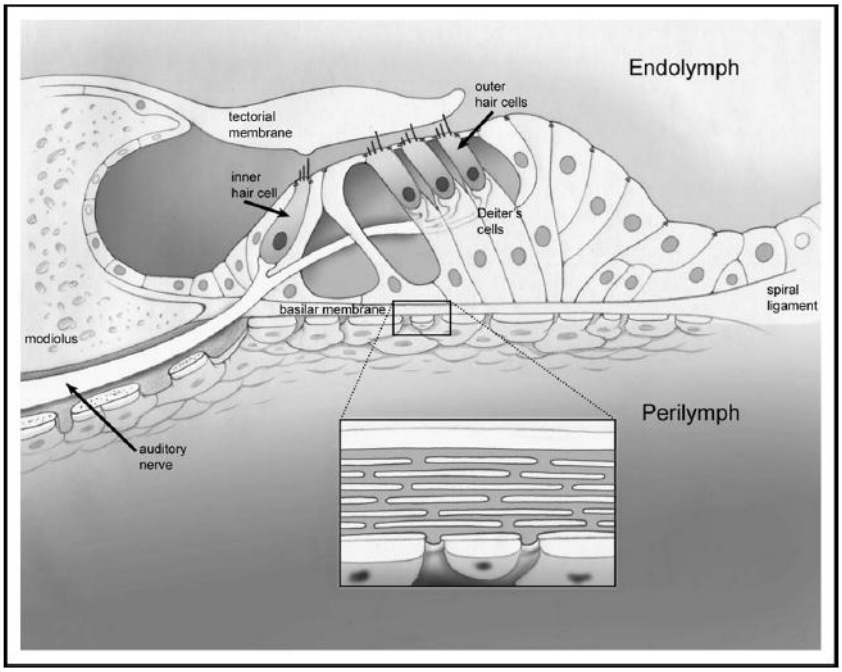

Figure 1.1: A diagram of the organ of Corti[1]. Note that the outer hair cells are rooted in the basilar membrane and their stereocilia extend to the tectorial membrane[1].

Over twenty percent of people in the United States over the age of twelve suffer from hearing loss in at least one ear[2]. Hearing loss by itself is obviously a severe disadvantage by itself, but the effect is compounded by the fact that hearing loss is often coupled with severe secondary effects if left untreated. In children, hearing impairment has been linked to developmental delay[3]. This delay is avoidable if a hearing deficit is detected and treated early on. Hearing impairment also tends to make it more difficult to pick out speech from background noise. In adults, hearing loss may lead to feelings of isolation[4] and depression[5].

[1]Reprinted with permission from J. S. Oghalai, "The cochlear amplifier: augmentation of the traveling wave within the inner ear," *Current opinion in otolaryngology & head and neck surgery*, vol. 12, pp. 431–438, Oct. 2004. Copyright 2004 Wolters Kluwer Health, Inc.

The most commonly used variety of hearing aid simply amplifies sound to compensate for the patient's decreased hearing threshold [6]. Some advanced hearing aids use noise filtering, audio compression, or selective amplification techniques in an attempt to amplify signal (speech, music, etc.) independent from background noise[7]. While these approaches do compensate for a patient's decreased hearing threshold, they fail to consider any of the frequency-related effects of diminished cochlea performance. In the course of transducing sound from a pressure wave into an electrochemical signal, different frequency components of sound are spread to distinct spatial locations along the cochlea. In a damaged cochlea, certain frequencies may resonate over a larger portion of the cochlea than normal leading to a decrease in frequency bandwidth and a degraded ability to discriminate between frequencies. The result is that speech may be more audible than with an amplification-only hearing aid, but not as intelligible as they would be to a healthy ear.

The entire organ of Corti vibrates up and down when interacting with sound in the form of a mechanical pressure wave. As the basilar membrane and tectorial membranes are coupled to the modiolus at different points (see Figure 1.1), a shearing motion between the two membranes occurs during the course of these vibrations. Outer hair cells are comprised of two major components: bodies rooted in the basilar membrane, and stereocilia which attach to the tectorial membrane. A shearing motion between the basilar and tectorial membranes will then, naturally, cause some deflection of the outer hair cells' stereocilia. Stereocilia extend and retract from the cell body in response to deflection in a process known as electromotility[8]. Outer hair cell electromotility provides an active force which opposes vibrational damping. As the basilar membrane changes in thickness and mass from its base to its apex, frequencies naturally resonate at different spatial positions on the cochlea. In effect, the cochlea is able to separate frequency components of an audio signal into different spatial locations and then separately amplify different frequency components depending on which outer hair cells are active. From these characteristics, it is logical that this behavior is commonly referred to as the cochlear amplifier[1]. Inner hair cells are responsible for the final transduction of a mechanical signal to an electrochemical signal which is transmitted through the auditory nerve to the brain. The inner hair cells have a similar composition to outer hair cells,

but differ in their location and their function. When stereocilia of the inner hair cells are deflected by the tectorial membrane, the inner hair cells release neurotransmitters which ultimately trigger action potentials in the auditory nerve. There does currently exist a model which predicts auditory nerve action potential activity resulting from an arbitrary sound wave[9]. However, this model does not currently include any functionality which simulates the response of a deaf or hearing-impaired cochlea.

Sensioneural hearing loss is commonly caused by some sort of damage to the outer hair cells[10]. As the location of outer cells along the cochlea dictate which frequencies they respond to, loss of or damage to these hair cells ultimately degrades the ability for the cochlea to amplify certain frequencies for transduction by the inner hair cells. Degraded amplification clearly results in elevated hearing thresholds which makes it more difficult to detect sound, but there are also several frequency-related effects which are visible in experiment and simulation that are not treated by common hearing aids.

Figure 1.2 shows the simulated comparison between the basilar membrane responses for active and passive cochlea. In the passive cochlea model, all outer hair cell function is suppressed to represent the behaviour of a dead cochlea lacking any active amplification mechanism. The passive cochlea's decreased response amplitude is obviously visible. Two frequency-related effects are also clearly present. First, the width of the response peak is much more broad in the passive cochlea response than in the active cochlea response[11]. Decreased peak sharpness results in decreased frequency selectivity, which may play a major role in speech intelligibility[12]. Second, the tip of the response peak appears to be shifted toward the base of the basilar membrane. This means the tone will be perceived as being of higher frequency than it actually is[13]. These frequency-related effects are even more evident when the audio signal input to the passive cochlea model is amplified such that the response peak is the same magnitude as that of the active cochlea as seen in Figure 1.3. This is the type of response expected to be caused by a sinusoidal signal amplified by a conventional hearing aid. The goal of this work is to devise a signal processing method which may compensate for these frequency-related consequences.

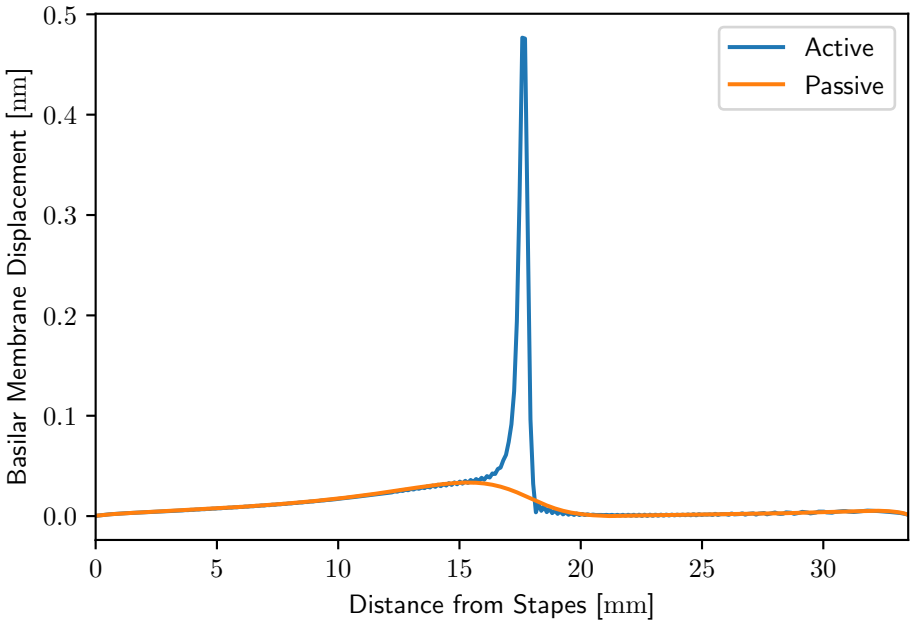

Figure 1.2: Simulated basilar membrane displacement of both the active and passive cochlea in response to a $4\,\mathrm{kHz}$ sine wave at approximately $40\,\mathrm{dB}$ SPL.

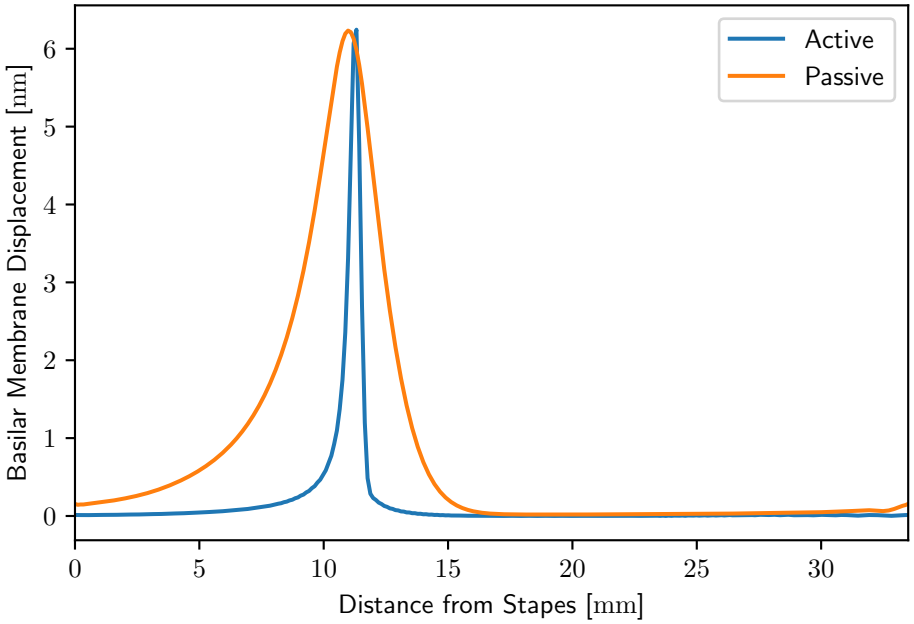

Figure 1.3: Active basilar membrane response compared to amplified passive basilar membrane response to a $4\,\mathrm{kHz}$ sine wave at $40\,\mathrm{dB}$ SPL.

2. THE COCHLEA MODEL

2.1 The Helmholtz Model

In the late nineteenth century, Helmholtz posed a conceptual model to describe the mechanism by which the cochlea is able to distinguish between frequencies in an audio signal comprised of many different frequencies[14]. He imagined the structure in the inner ear to be comprised of tensioned strands with differing natural frequencies having some degree of sympathetic vibration between the strands similar to the construction of a piano.

This idea can be related to the damped harmonic oscillator:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + h\frac{\mathrm{d}x}{\mathrm{d}t} + kx = f(t) \tag{2.1}$$

Above, m is the oscillator mass, h is a damping coefficient, k is a spring constant, and $f(x)$ is a driving force. The damped driven harmonic oscillator of course has a resonance at $w_0 = \sqrt{k/m - h^2/4m^2}$. Clearly, the resonant frequency is dependent on the mass, damping coefficient, and spring constant of the oscillator. If we consider the basilar membrane to be a discretized collection of oscillators rather than a continuous membrane we have a set of differential equations:

$$m_i\frac{\mathrm{d}^2x_i}{\mathrm{d}t^2} + h_i\frac{\mathrm{d}x_i}{\mathrm{d}t} + k_ix_i = f_i(t) \qquad i = 1, \ldots, N \tag{2.2}$$

The index i represents a section of the Basilar membrane. So, the coefficients m_i and k_i represent the mass and stiffness of a section of basilar membrane while h_i represents the viscosity of the fluid surrounding a section of the basilar membrane and $f_i(t)$ represents the driving force transmitted to a section of the basilar membrane. While this model does describe frequency localization on the cochlea, this model requires some expansion through the addition of several terms which will make the behaviour of the model more similar to that of a real cochlea.

2.2 Expanding upon the Helmholtz Model

There are three major factors described by Mammano & Nobili[15] which are missing from the Helmholtz cochlea model.

The first term they describe is the shear viscosity term. This term represents force on an oscillator due to differing velocity of its neighbors and has the form:

$$s_i^+ \left(\frac{\mathrm{d}x_{i+1}}{\mathrm{d}t} - \frac{\mathrm{d}x_i}{\mathrm{d}t} \right) - s_i^- \left(\frac{\mathrm{d}x_{i-1}}{\mathrm{d}t} - \frac{\mathrm{d}x_i}{\mathrm{d}t} \right) \tag{2.3}$$

They further consider the simplification $s_i = s_i^+ = s_i^-$ meaning that an oscillator to the left of oscillator i imparts the same force due to a difference in velocity as the oscillator to the right of oscillator i. This term then becomes:

$$s_i \left(2\frac{\mathrm{d}x_i}{\mathrm{d}t} - \frac{\mathrm{d}x_{i-1}}{\mathrm{d}t} - \frac{\mathrm{d}x_{i+1}}{\mathrm{d}t} \right) \tag{2.4}$$

The second term has to do with the connection between oscillators. In the Helmholtz model, there is no definite connection between oscillators meaning the response will be a stationary wave. As the basilar membrane is a continuous entity driven from one end, it's more reasonable to assume that the response should be a traveling wave. This is conceptually similar to the waveform produced by shaking one end of a rope. To account for this, Mammano & Nobili add a series of terms accounting for the hydrodynamic coupling between oscillators:

$$-\sum_{j=1}^{N} G_i^j \frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} \tag{2.5}$$

This term represents the change in acceleration of oscillator i due to motion of all other oscillators in the system. These other oscillators are indexed with j. The coefficients G_i^j above come from a matrix of Green's functions which describe the hydrodynamic properties of the cochlea.

To this point the model only describes the behaviour of the passive cochlea while entirely neglecting the cochlear motor effect which is responsible for the amplification and frequency discrim-

ination that distinguishes mammalian hearing. This leads to the third and possibly most important term:

$$- U_i(y_i) \tag{2.6}$$

The function U_i is an undamping term which accounts for opposition to the damping force provided by the outer hair cells. Its argument, y_i, is the outer hair cell stereocilia displacement. When driven by a very high amplitude sound wave, the cochlear amplifier in a healthy ear will saturate and the response of the active and passive cochlea will be nearly the same. To account for this, the outer hair cell stereocilia displacment argument y_i follows a sigmoidal shape. In practice, U_i takes the form of a constant multiplied by the outer hair cell stereocilia displacement for each oscillator in the system: $U_i(y_i) = u_i y_i$. The constant defines the hearing ability of the cochlea. As there is one constant term u_i corresponding with each osccilator, these terms can be set to zero to model a fully-passive cochlea, or they may be set to model hearing damage over a pre-determined range of frequencies.

Combining the Helmholtz model in Equation 2.2 with the additional terms in Equations 2.4, 2.5, and 2.6 gives us the final functional form for our model of the basilar membrane dynamics:

$$\sum_{j=1}^{N} \left(G_i^j + m_i \delta_j^i\right) \frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} + h_i \frac{\mathrm{d}x_i}{\mathrm{d}t} + U_i(y_i) + s_i \left(2\frac{\mathrm{d}x_i}{\mathrm{d}t} - \frac{\mathrm{d}x_{i-1}}{\mathrm{d}t} - \frac{\mathrm{d}x_{i+1}}{\mathrm{d}t}\right) + k_i x = -G_i a_s(t) \tag{2.7}$$

The force term $f_i(t)$ has been replaced by the product of a Green's function (G_i) representing the hydrodynamic coupling between the stapes and the basilar membrane and the acceleration of the stapes ($a_s(t)$) due to a sound pressure wave. Also note that the combination of the acceleration terms requires the addition of the Kronecker delta δ_j^i which is equal to 1 if $i = j$ and 0 if $i \neq j$.

This model was implemented in MATLAB by Mammano & Nobili and is available in its original form from their website. For ease of code execution, integration with optimization utilities, and data analysis, I have ported the original MATLAB model to Python.

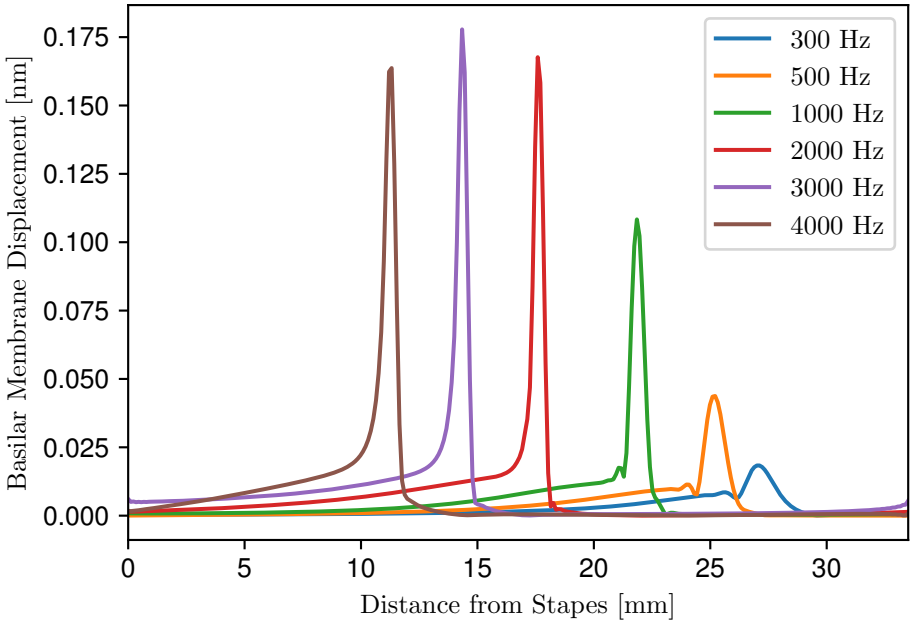

Figure 2.1: Signals with different frequencies respond at different locations on the cochlea. Note that the response amplitude changes with frequency due to nonlinear undamping from the outer hair cells in the active cochlea.

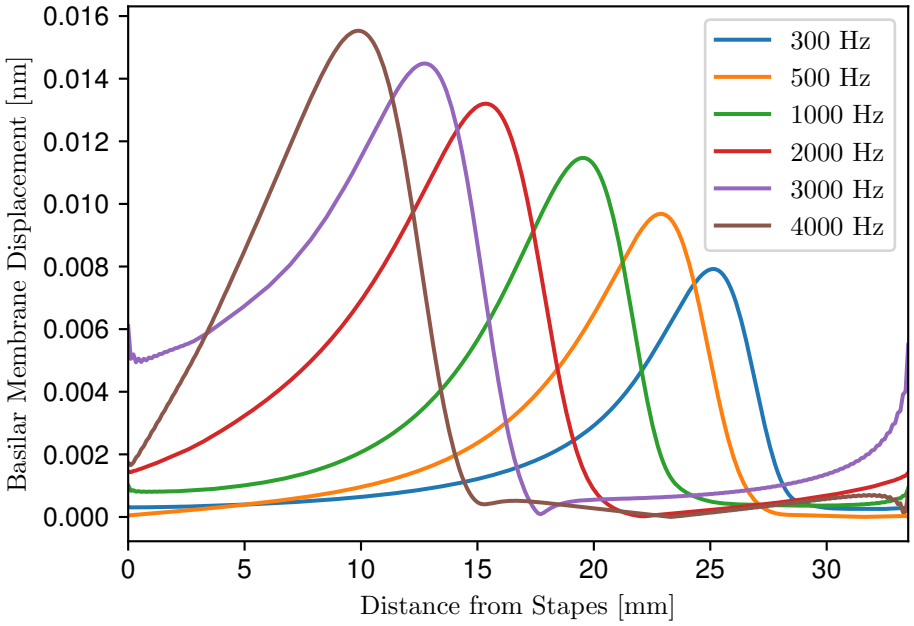

Figure 2.2: The passive cochlea response is much broader and lower amplitude than the active cochlea response. Notice the change in displacement units. Response peaks are also shifted slightly toward the stapes when compared to the active cochlea responses in Figure 2.1.

3. FILTERING AND OPTIMIZATION

The direct goal of this work is to alter an audio signal such that the passive cochlear response like that seen in Figure 2.2 more closely resembles the response of the active cochlea as in Figure 2.1. Comparing these two figures reveals three striking deficits of the passive cochlea. First is the decrease in response amplitude. This is commonly treated using conventional hearing aids which simply amplify the sound. The second is that the response peak is shifted nearer to the stapes. This effect is simply compensated for by simply shifting the central frequency of the tone to lower frequency. The third effect is the broadening of the response peak on the passive cochlea. This effect is the focus of this work.

The idea is to take advantage of how tones interact in the cochlea[16] to mask portions of the passive cochlear response in order to sharpen the peak and improve the frequency selectivity of a cochlea with diminished outer hair cell activity.

This chapter will discuss several possibilities for filter design, optimization methods, and similarity metrics.

3.1 Construction of a New Audio Signal

The most obvious choice of audio filter following from the concepts introduced earlier is the addition of signals to the central pure tone:

$$S_f = a_0 \sin\left(2\pi[f_0 + \Delta f]t\right) + \sum_{i=1}^{N} a_i \sin(2\pi f_i t + \phi_i) \tag{3.1}$$

Careful selection of the parameters a_0 and Δf can correct for response amplitude and location, but we rely on the summation on the right hand side of this equation for shaping the response peak. This method is unfortunately susceptible to the curse of dimensionality as the number of parameters to optimize grows as $3(\mathrm{N}+1)$. In addition to this, it is not possible to optimize the frequency parameters f_i independent of the phase parameters ϕ_i or amplitude parameters a_i as frequency, amplitude, and phase all influence the behaviour of wave additivity. It is worth investigating other

methods for signal construction.

3.1.1 Representation of the Audio Signal as a Linear Combination of Basis Functions

It is well known that an arbitrary vector in a vector space can be expressed as the linear combination of orthonormal basis vectors. From this we can consider that if the optimal waveform exists it may be constructed with a linear combination of orthonormal basis functions.

This is very similar to the idea of Fourier expansion in which an arbitrary function can be represented by a linear combination of sines and cosines[17]. In this case, we'd like to represent a linear combination of sinusoidal waves with arbitrary phase (ϕ_i in Equation 3.1). A sine wave of the form $\sin(2\pi ft + \phi)$ can be represented as a combination of a sine and cosine without phase terms: $a_1 \sin(2\pi ft) + a_2 \cos(2\pi ft)$. An example of this can be seen in Figure 3.1. Constructing the signal this way allows us to build the matrix of basis functions (B in Equation 3.2) once at the beginning of the optimization. Constructing a new signal at each iteration of the optimization only requires a simple dot product between the B matrix and the vector of optimized amplitudes, a. This form is equivalent to, but computationally cheaper than constructing the signal using the form in Equation 3.1.

$$S_f = B \cdot a = \begin{bmatrix} \sin(2\pi f_1 t) \\ \cos(2\pi f_1 t) \\ \sin(2\pi f_2 t) \\ \cos(2\pi f_2 t) \\ \vdots \\ \sin(2\pi f_{n-1} t) \\ \cos(2\pi f_n t) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix} \tag{3.2}$$

Both sine and cosine terms for each frequency are present in the basis set B, so we have $2\mathrm{N}$ variables for which we must find optimal values. For this method to be most successful, the frequencies present in the optimal signal must be present in the basis set.

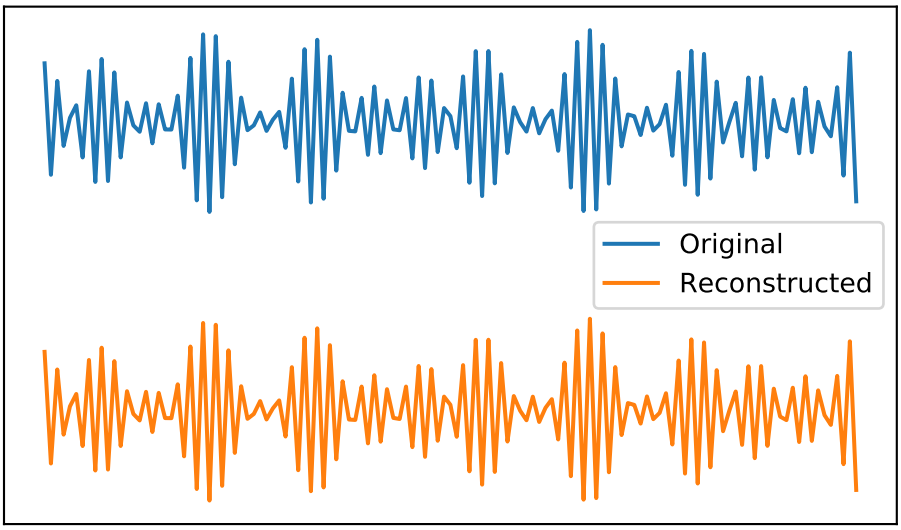

Figure 3.1: An arbitrary sinusoid reproduced by sine and cosine basis functions. The top signal is $S_0 = 2.86\sin(2\pi 20t + \pi/16) + 3.5\sin(2\pi 18t + \pi/2)$ and is represented on the lower half of the figure by a linear combination of basis functions: $S_f = 2.8061\sin(2\pi 20t) + 0.5646\cos(2\pi 20t) + 0.0016\sin(2\pi 18t) + 3.4997\cos(2\pi 18t)$. The reconstructed signal is shifted in this figure for clarity.

3.2 Selecting an Objective Function and Optimization Method

Choosing the best parameters for either of the filters in Equations 3.1 or 3.2 can be written as an optimization problem with box constraints as in Equation 3.3. Box constraints simply mean that each member a_i of the vector a must be larger than some lower bound l but smaller than some upper bound u.

$$\begin{aligned} \text{minimize} \quad & \|A(S_0) - P(S_f)\|_2^2 \\ \text{subject to} \quad & S_0 = a_0 \sin(2\pi f_0 t), \\ & S_f = B \cdot a, \\ & l \preceq a \preceq u \end{aligned} \tag{3.3}$$

In the above equation, $A(S_0)$ is the active cochlea response to the unfiltered signal, and $P(S_f)$

is the passive cochlea response to the filtered signal. It should be noted that this method may be applied to the dynamics of any structure in the cochlea. In this specific case, we're using the displacement of the basilar membrane as a surrogate to evaluate the hearing capability of the cochlea. The quantities $A(S_0)$ and $P(S_f)$ are matrices containing values for basilar membrane displacement as a function of both time and distance along the membrane. The sum of the square difference between the responses is selected as an objective function as ideally the objective function will be zero when the responses are the same with increasing penalty as the responses become more distant.

Gradient descent, stochastic gradient descent, and other popular optimization methods generally require information about the objective function's first or second derivative with respect to the decision variables and/or a large number of function evaluations. As the active response and passive response are both fairly computationally expensive, performing a large number of objective function evaluations in the course of optimization is not ideal. Additionally, it may be very difficult to determine the gradient of the objective function with respect to filter parameters, especially if the signal is generated as in Equation 3.1. Finally, we cannot make any assumptions about the objective function's topology with respect to the filter parameters. This last point rules out the possibility of using a local search method as these methods are susceptible to getting stuck in saddle points or local minima, and these methods also cannot guarantee that the found optimal value is globally optimal unless the objective function is convex[18] with respect to the decision variables.

Based on the above considerations, a gradient-free optimization technique geared toward high-dimensional optimization problems with expensive objective function evaluation is desirable. With this type of optimization problem, it may be prudent to consider evaluating a less-expensive surrogate to the objective function. The Metric Stochastic Resopnse Surface (MSRS) response technique operates much this way[19]. Several points or "nodes" are chosen stochastically throughout the pre-defined feasible space. The objective function is evaluated at each of these nodes, and the resulting values form the surface response of the objective function. A radial basis function (RBF) interpolation scheme is used to predict the behavior of the objective function between nodes to

identify locations of new possible minima. When a new suspected minima is identified, a local optimization may be executed in its vicinity in an effort to improve upon the local minima. Optimization terminates if the objective function is acceptably close to the target value, or after a set number of objective function evaluations. This optimization technique is implemented in the `rbfopt` python package[20]. The `rbfopt` package also implements the Gutmann RBF optimization algorithm[21]. RBFOpt includes other functionality such as the automated selection of radial basis function (Gaussian, multiquadric, thin plate spline), automatic selection of radial basis function shape parameter, and functionality to avoid non-optimal local minima by stochastically re-sampling the optimization space after a set interval.

4. RESULTS AND CONCLUSIONS

4.1 Results

All results in this section use a $2\,\mathrm{kHz}$ sinusoidal tone with an amplitude of $40\,\mathrm{dB\ SPL}$ which corresponds to approximately $0.0494\,\mathrm{m/s^2}$ stapes acceleration (see Appendix A.1 for conversion). An input signal fitting this description is shown in the top half of Figure 4.1. Before being input to the model, the amplitude must be converted from stapes acceleration in $\mathrm{m/s^2}$ to the model input in β/s. The β unit is a derived unit where $1\,\beta$ corresponds to the length of the cochlea, approximately $33.5\,\mathrm{mm}$. This conversion is described in Appendix A.2. The initial condition of the model assumes that the stapes is at rest. If the input signal drives an instantaneous change in stapes acceleration, a severe artifact similar to switch noise or the Gibbs phenomenon appears in the response. This is sometimes referred to as "spectral splatter" because sizeable responses are visible in regions of the cochlea that shouldn't be driven by the input tone[22]. To avoid this artifact, a gradual ramping function from zero to one is multiplied with the input signal. For convenience, we use the hyperbolic tangent function because it is the proper shape and simple to adjust. An amplitude-ramped signal is shown in the bottom half of Figure 4.1. The active response to this input signal is displayed in Figure 4.2. Signals used for these trials have a duration of $25\,\mathrm{ms}$. Although the optimization strategy utilized here is designed to require a minimal number of objective function evaluations, the computational expense of the objective function does still have a large impact on the total optimization time. The time required to compute the basilar membrane response increases substantially as the input signal length increases. Increasing the length of the input signal also increases the optimization problem's complexity. The objective function condenses a fairly large amount of information about the difference between the healthy cochlea response to the unfiltered signal and the damaged cochlea response to a filtered signal into a single scalar value. Extending the duration of the input signal means an increasingly large amount of information is condensed into this value. At this stage, it is sufficient to find a solution

that works for a shorter time duration and explore methods to identify a solution for longer input signals in the future.

As discussed in §2.2, the undamping term in the cochlea model can be modified at will to model different types of hearing damage. The undamping parameter used in the first set of results is a 25% loss in cochlear amplifier activity between $1900\,\mathrm{Hz}$ and $2100\,\mathrm{Hz}$, and can be seen in Figure 4.7. The second set of results uses uniform hearing loss which is increases in severity until no improvement in basilar membrane response can be produced through signal optimization.

4.1.1 Notch Damage

Optimization takes place in two stages. First, a frequency and phase are found such that the passive response has the same amplitude and peak location as the active response. In this case, the original frequency and amplitude are $f_0 = 2\,\mathrm{kHz}$ and $a_0 = 40\,\mathrm{dB\ SPL}$ as in Equation 4.1. After optimization, the selected frequency and phase are $f_{p0} \approx 1977\,\mathrm{Hz}$ and $a_{p0} \approx 50.2\,\mathrm{dB\ SPL}$ as in Equation 4.2. The result of the first optimization stage is shown in Figure 4.4.

Using the first stage optimization result (Equation 4.2) as a starting point, the signal used in the second stage of optimization will be constructed as in Equation 4.3. As before, B is a matrix containing the basis set of signals, and $\vec{a}$ are coefficients which are to be optimized.

$$S_0 = a_0 \sin\left(2\pi f_0 t\right) \tag{4.1}$$

$$S_{p0} = a_{p0} \sin\left(2\pi f_{p0} t\right) \tag{4.2}$$

$$S_f = S_{p0} + B \cdot \vec{a} \tag{4.3}$$

The goal is to narrow the peak of the damaged cochlea response so that it looks more like the active response peak. There are two considerations that can be made by comparing the two curves in Figure 4.4. The first is that the left (high-frequency) side of the peak shows more difference than

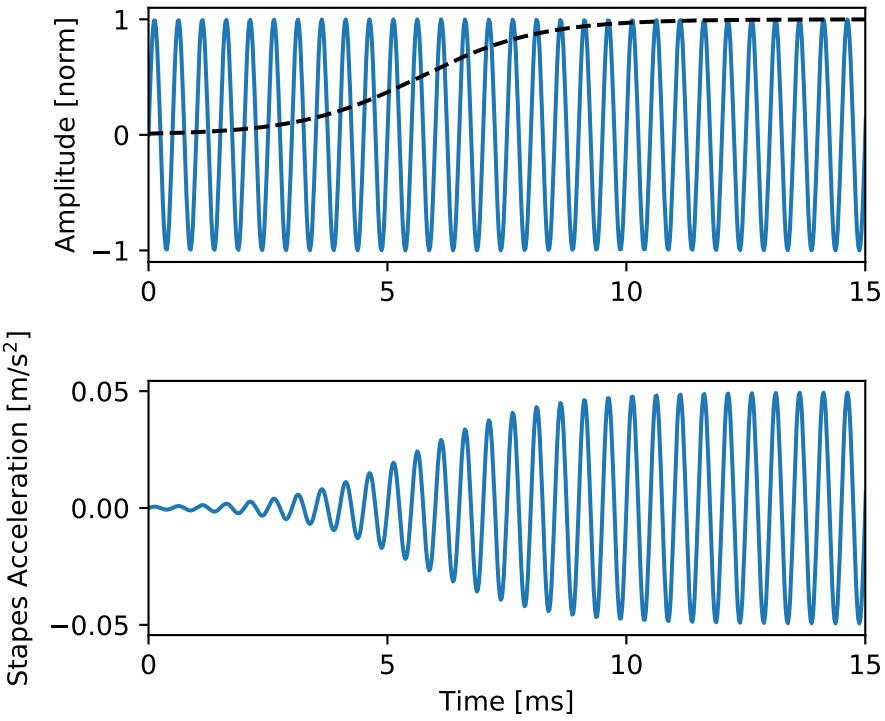

Figure 4.1: Input audio signal with ramped amplitude. The top half of this figure shows an unmodified sine wave with the amplitude envelope function shown by the dashed line. In the bottom half, the sine wave has been multiplied with the envelope function. The signal in the bottom half of this figure may be scaled by a constant to set its amplitude and then it will be ready for use as a model input.

the right (low-frequency) side. This suggests that we should put more frequencies in the basis set above the central frequency than below. The second is that frequencies near the central frequency may provide a larger impact on peak width than those further away. It is also important to include as few frequencies in the basis set as possible while still achieving a satisfactory result. The number of optimization values scales as $2n$ where n is the number of frequencies in the basis set. For these reasons, the frequency distribution is made up of two geometrically-spaced subdistributions: one below the central frequency, and one above. Creating these subdistributions separately gives the freedom to select how many frequencies are added to the basis set on each side of the central frequency, and what degree of geometric stretching is applied on each side. Figure 4.5 shows a visualization of the final frequency distribution.

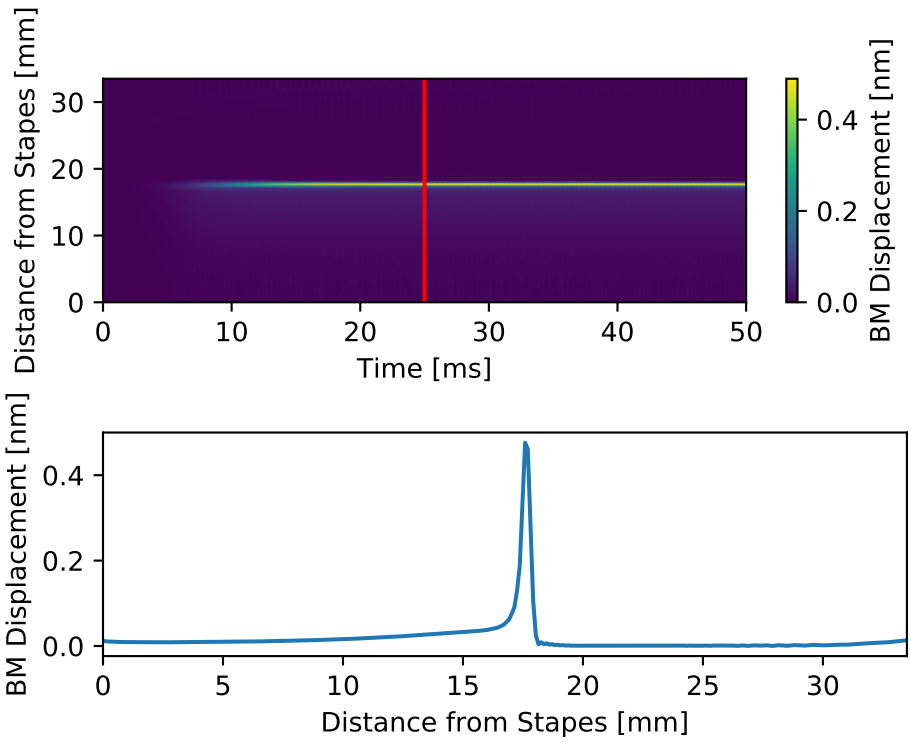

Figure 4.2: Active response to the signal in Figure 4.1. The top half shows the response in both space and time, and the bottom half is a spatial plot of the response amplitude as a function of length along the basilar membrane taken at the red vertical line in the top half.

Although the ℓ_2-norm objective function is discussed earlier, the Huber loss function, Equation 4.4 was used for final optimization.

$$H_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } a \leq \delta \\ \delta\left(|a| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases} \tag{4.4}$$

An example of the Huber function is shown in Figure 4.6. This function is popular in regression as it is less sensitive to outliers than objective functions using ℓ_1 or ℓ_2 norms. The Huber function was applied to each "pixel" in 2D data similar to that shown in Figure 4.2.

The final response result is seen in Figure 4.8. This plot shows a time slice from the 2D response data in the region of improved peak width indicated by Figure 4.9. The measure of peak width used here is Q10dB which is the width of the peak at $-10\,\mathrm{dB}$ with respect to the maximum

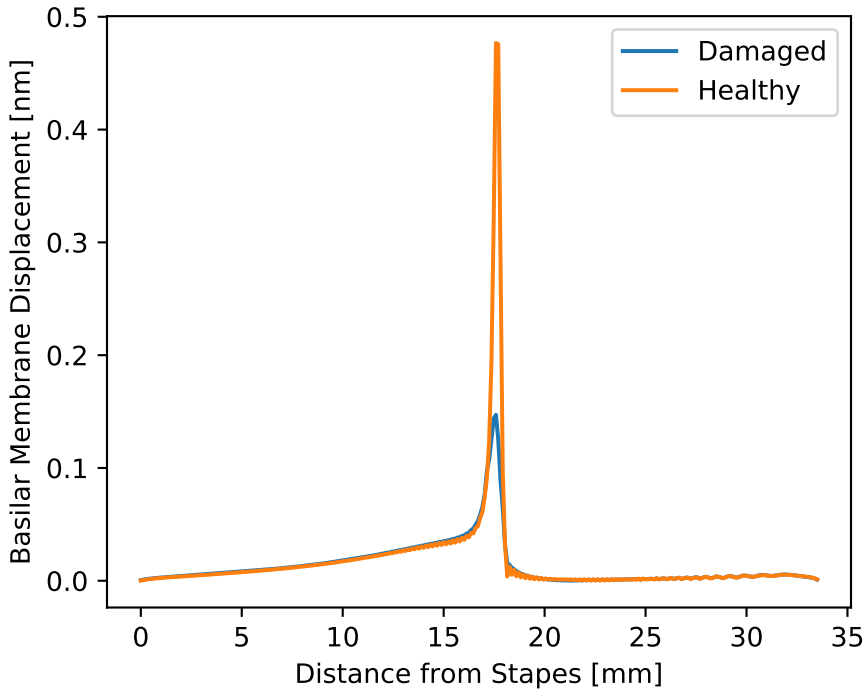

Figure 4.3: Unfiltered damaged cochlea response to the signal in Figure 4.1. The healthy response is also displayed for comparison.

peak value – about 0.316 times the maximum value. This point is indicated by the horizontal red line in Figure 4.8. An example of the optimized signal is plotted in Figure 4.10. Of course there is some temporal delay from signal to response, but the change in signal amplitude and shape from $20\,\mathrm{ms}$ to $30\,\mathrm{ms}$ seems to correspond with the improvement in peak width in Figure 4.9. The method used here is most effective for short time durations, these results do suggest that exploitation of this phenomenon is feasible to improve frequency selectivity of the damaged cochlea. This technique may provide more improvement for more realistic, time-varying audio signals such as speech rather than the artificial sine waves used here.

4.1.2 Uniform Damage

The results in this section use a uniformly decreased undamping coefficient to simulate the case of uniform outer hair cell damage. The undamping coefficients used for these tests are shown in

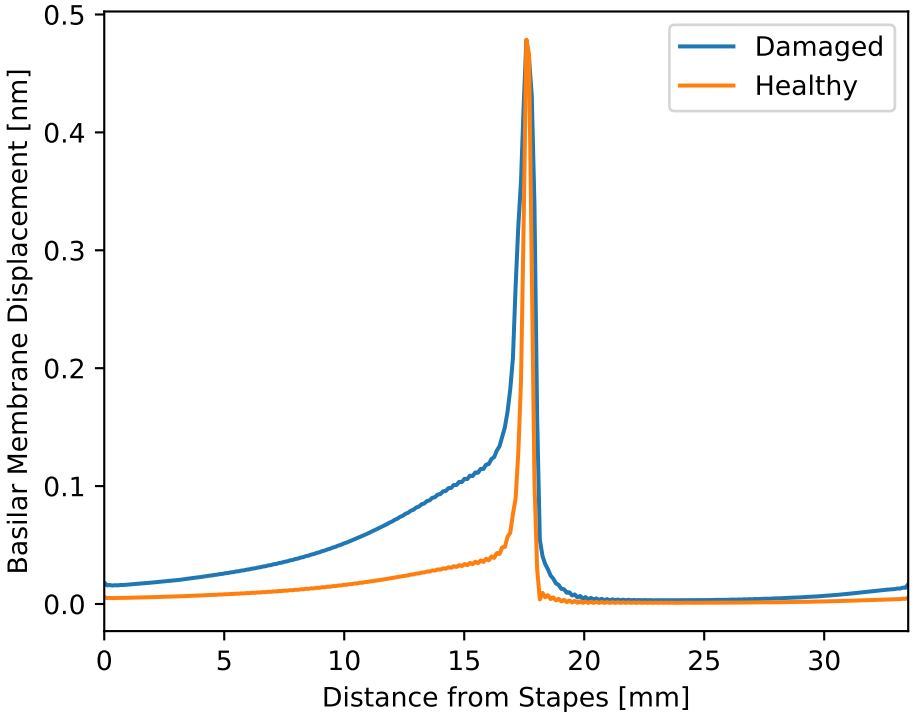

Figure 4.4: Result of the first optimization stage. The active and damaged responses now have the same amplitude, and the peaks are in the same location.

Figure 4.11. Otherwise, the optimization procedure is identical to that in the above section. Optimizations for the uniform hearing damage case started with 40% hearing damage and decreased in 10% increments until the response improvement from optimization is nearly non-existent. Figures 4.12, 4.13, and 4.14 show a summary of the basilar membrane response for a 40%, 50%, and 60% decrease in undamping coefficient. It is clear from comparing these three figures that the duration for which this optimization scheme is able to improve the Q10dB response peak width decreases as the degree of hearing damage increases. This does not indicate that no improvement can be expected with higher degrees of hearing loss, but that significant changes to the signal generation technique may be required to see basilar membrane response improvement in these types of hearing loss.

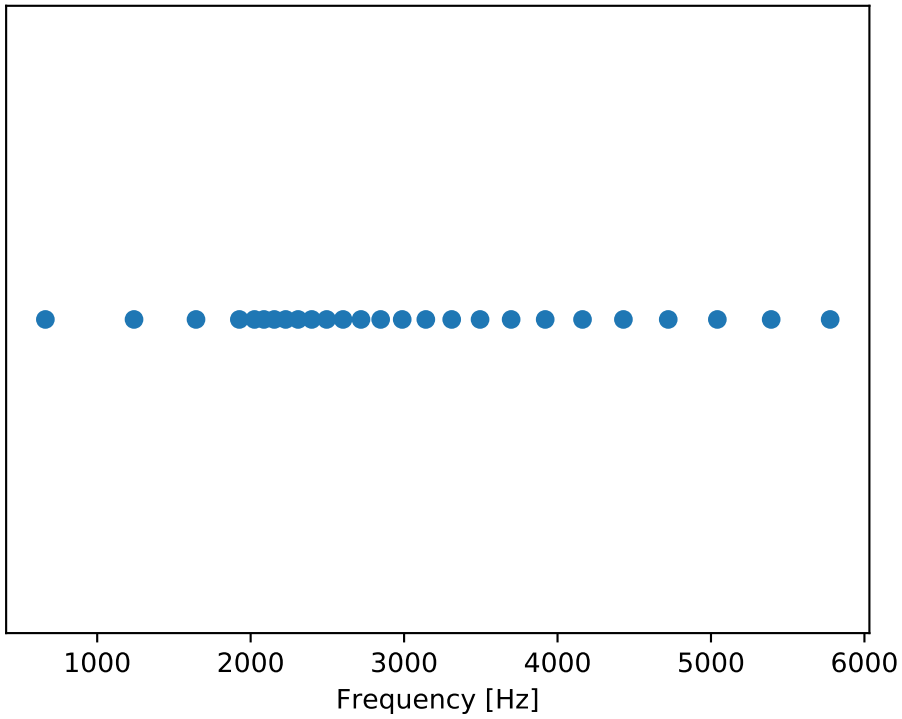

Figure 4.5: An example of the frequency distribution used to construct the basis set. The basis set used for optimization is comprised of 104 frequencies. This plot shows a subset of those frequencies to visualise their distribution.

4.2 Future Work

While the work presented here does show the feasibility of the concept of improving frequency selectivity in the damaged cochlea by signal processing, there are certainly opportunities that should be exploited in the future which may give better results.

One possibility is to more fully explore the possibility of modifying the auditory nerve activity model[9] to predict the behaviour as the passive cochlea. It is certainly possible that the auditory nerve action potential pulse trains output by this model may serve as a better indicator of audio perception. It is a distinct possibility that this filtering technique works quite well, but that basilar membrane displacement as modeled here is not the best surrogate to determine sound perception.

The presented method optimizes only for the amplitude parameters in Equation 3.2 while selecting frequencies for the basis set that seemed to give the best results. It may be possible to

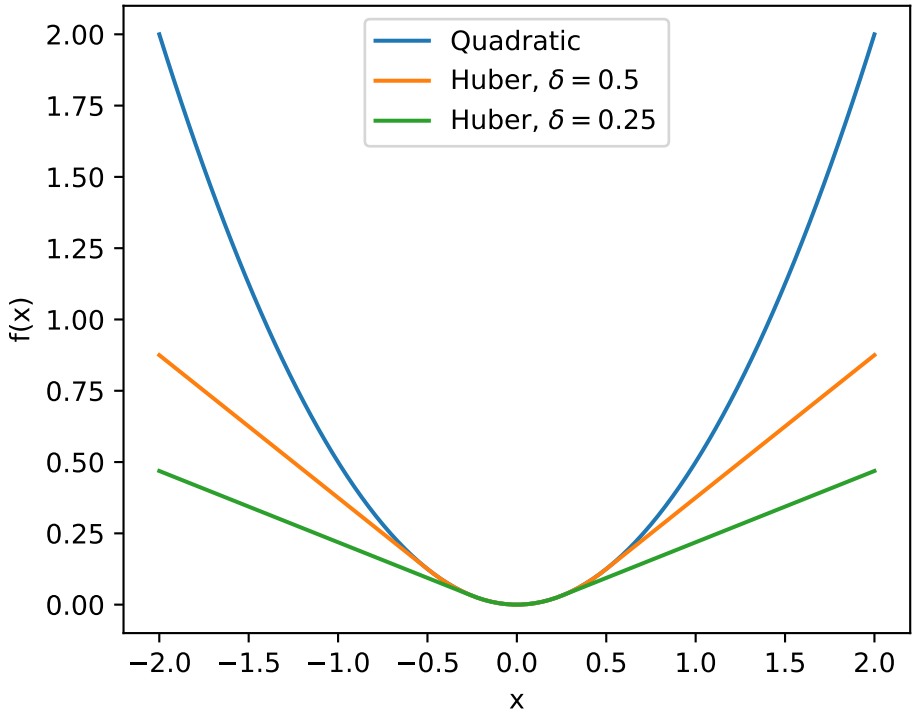

Figure 4.6: The Huber function plotted for two values of δ along with a quadratic function for comparison. The Huber function is strongly convex on the interval $\{-\delta, \delta\}$ and linear outside this interval. The location at which the function changes from convex to affine as well as the slope of the linear section is controlled by choice of δ.

devise a multi-stage optimization technique which has the ability to select these frequencies and their amplitude parameters. However, this would add decision variables and complexity to an already large and complex optimization problem. It may make more sense to apply a deep learning method which may be able to directly generate the optimal signal. Signals generated in the method I've discussed have quite a bit of flexibility, but this method is far from having the power to generate a completely arbitrary audio signal. Generating an arbitrary audio signal (i.e. optimizing the signal amplitude at each time point) is not feasible as the type of optimization algorithm may be unstable or unable to converge with such high dimensionality. Neural networks have the power to handle high-dimensional problems such as this, and deep learning should be explored as the next step to progress this work. It may be possible to adapt a network designed for speech synthesis such as WaveNet[23] to generate the audio signal with a custom loss function which evaluates the

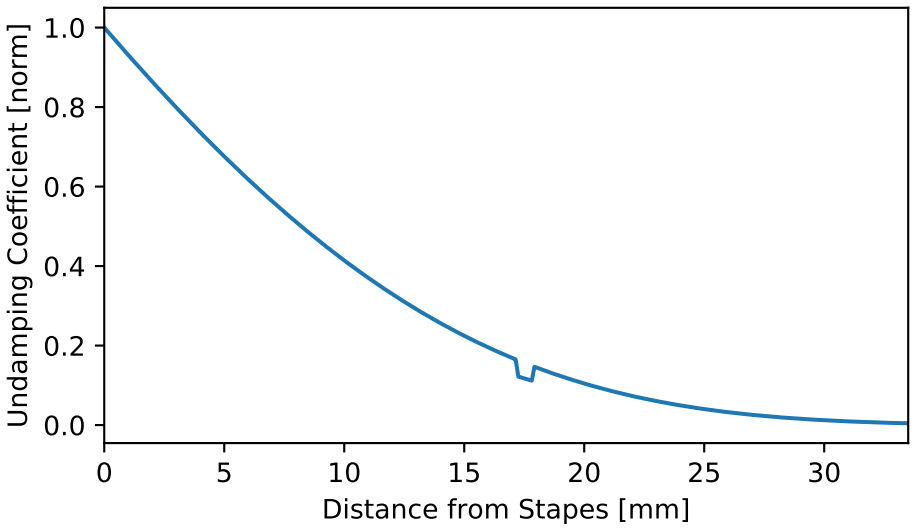

Figure 4.7: Undamping coefficient constructed to simulate notch-style hearing damage. The undamping coefficient is reduced by 25% between 1900 Hz and 2100 Hz.

passive cochlea model's response to the generated signal and compares it to the active response. Whichever way forward is chosen, this effort would benefit most by improving the generality of the signal generation algorithm.

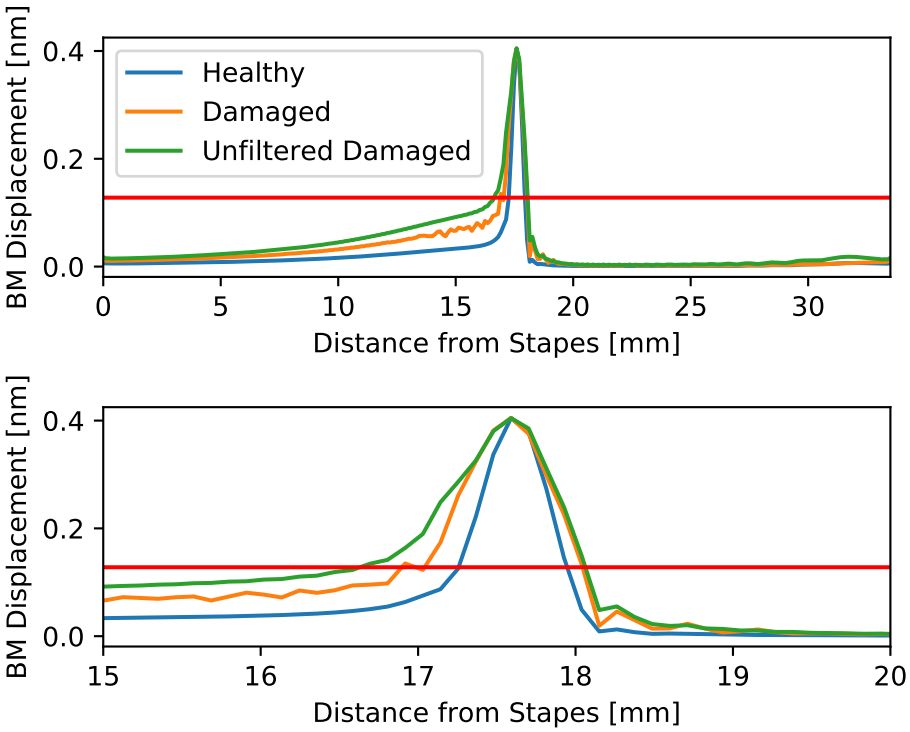

Figure 4.8: Result of the second optimization stage. This is a single time slice taken at $\approx 15\,\mathrm{ms}$. The top half of this figure shows the Basilar Membrane displacement for a single time slice. The bottom half is an expansion of the top half to show the peak shape improvement in more detail. The horizontal red line in both portions of this figure indicate the point at which the Q10dB peak width is calculated.

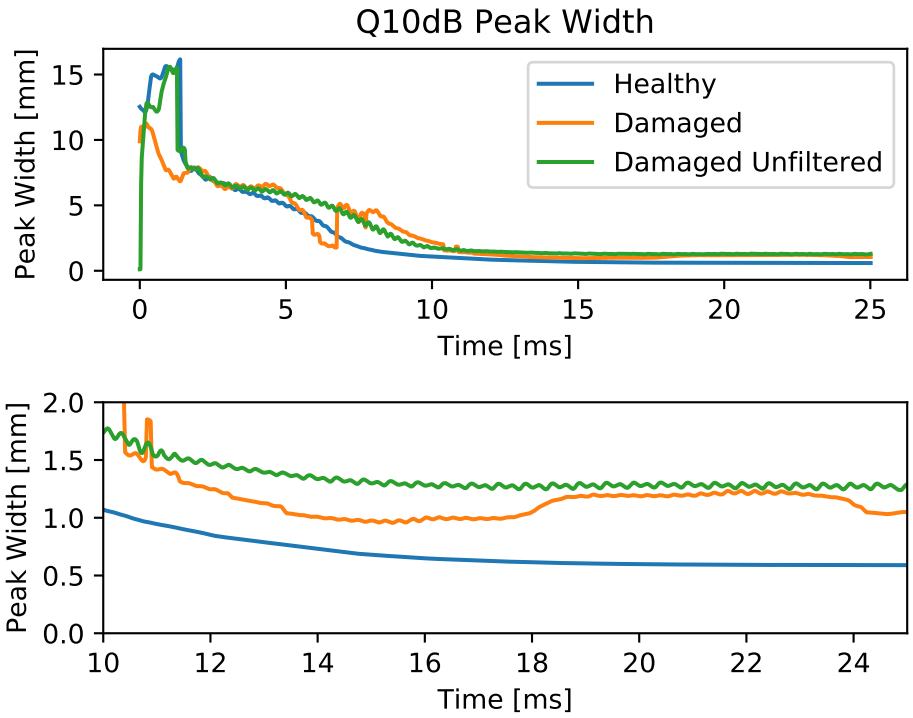

Figure 4.9: Q10dB Peak Width Comparison between the healthy, damaged, and unfiltered damaged responses. The top shows the peak width as a function of time for the entire signal duration. Optimization begins at $15\,\mathrm{ms}$. The response before $10\,\mathrm{ms}$ shows the period between the beginning of the signal and when the response approaches a "steady state" behavior. The bottom shows the region in which measurable peak width improvement can be seen for a duration of approximately $14\,\mathrm{ms}$.

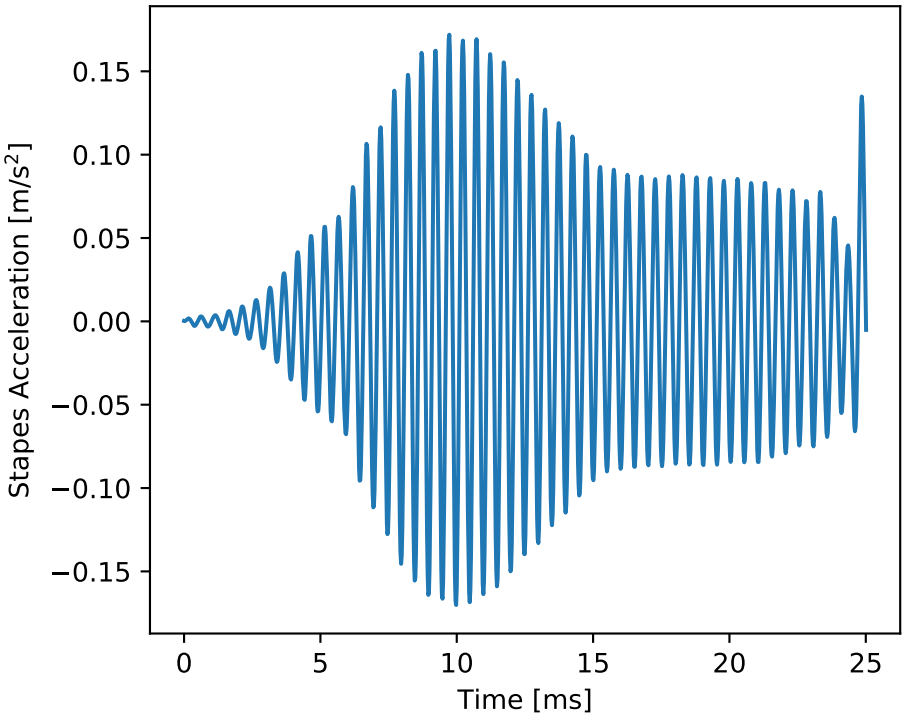

Figure 4.10: An example of a signal resulting from optimization. This signal was used to generate the response shown in Figure 4.8. Although the optimization is only carried out over a period from $15\,\mathrm{ms}$ to $19\,\mathrm{ms}$, the final signal shows a lower intensity from $15\,\mathrm{ms}$ to the end of the signal at $25\,\mathrm{ms}$. This region of lower intensity corresponds to a region of improved peak shape as indicated by Figure 4.9.

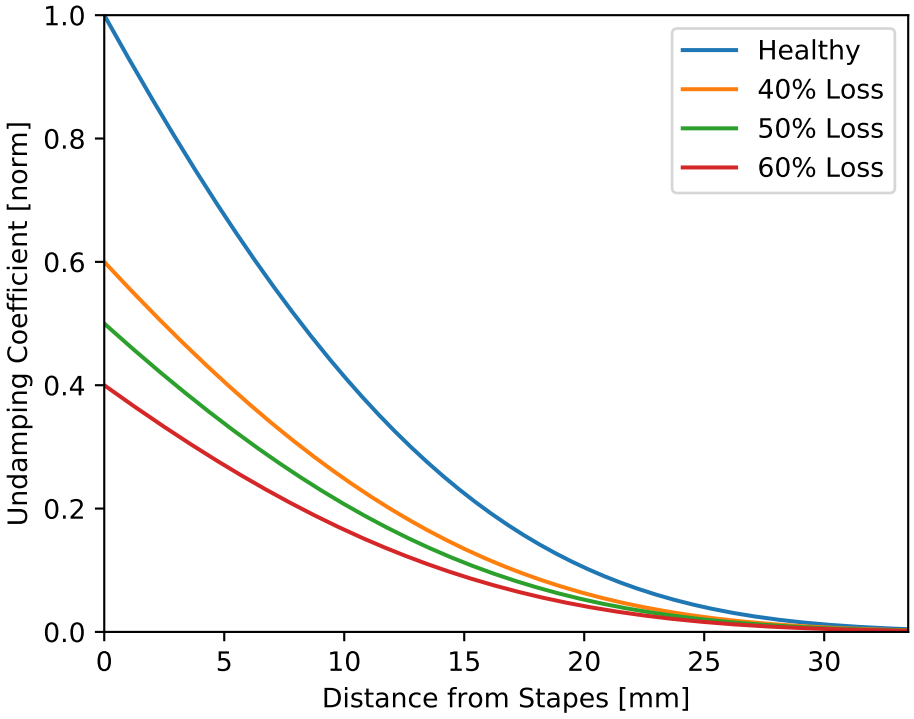

Figure 4.11: Undamping coefficients used for the uniform hearing damage cases.

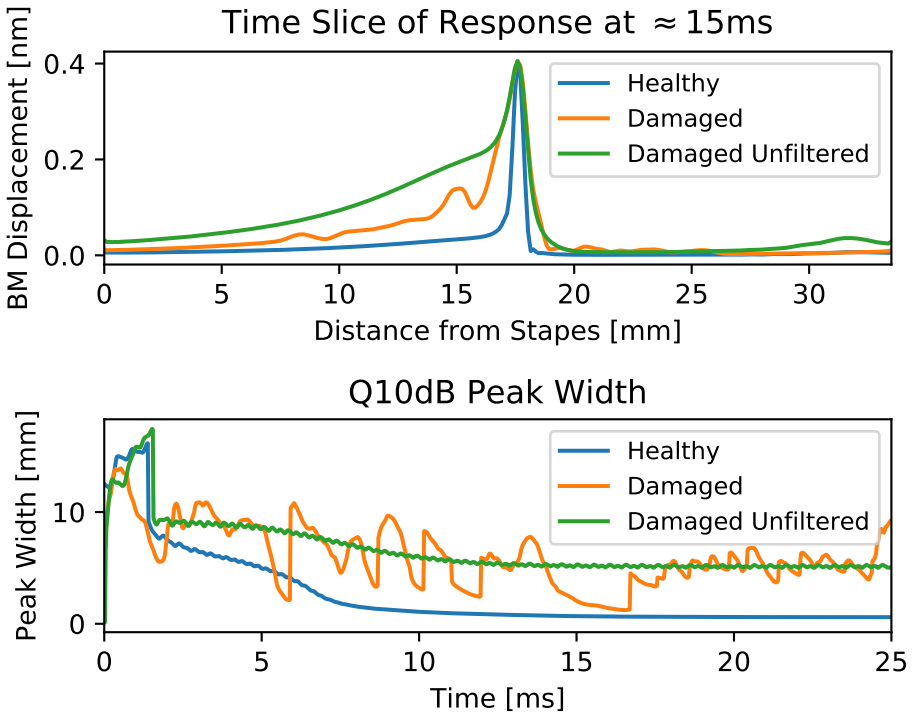

Figure 4.12: Optimization result for 40% uniform hearing damage.

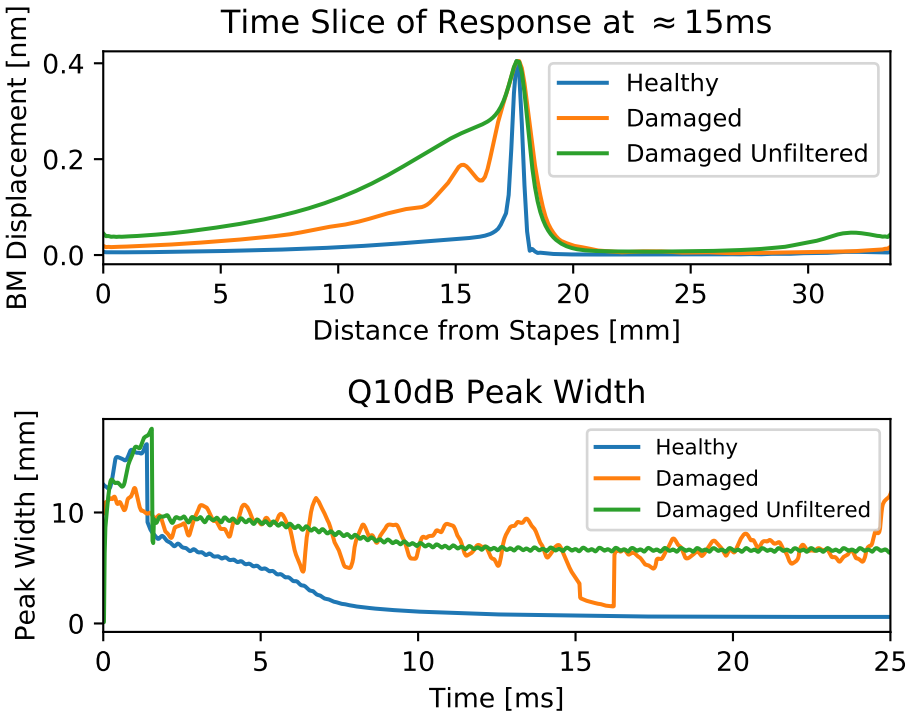

Figure 4.13: Optimization result for 50% uniform hearing damage.

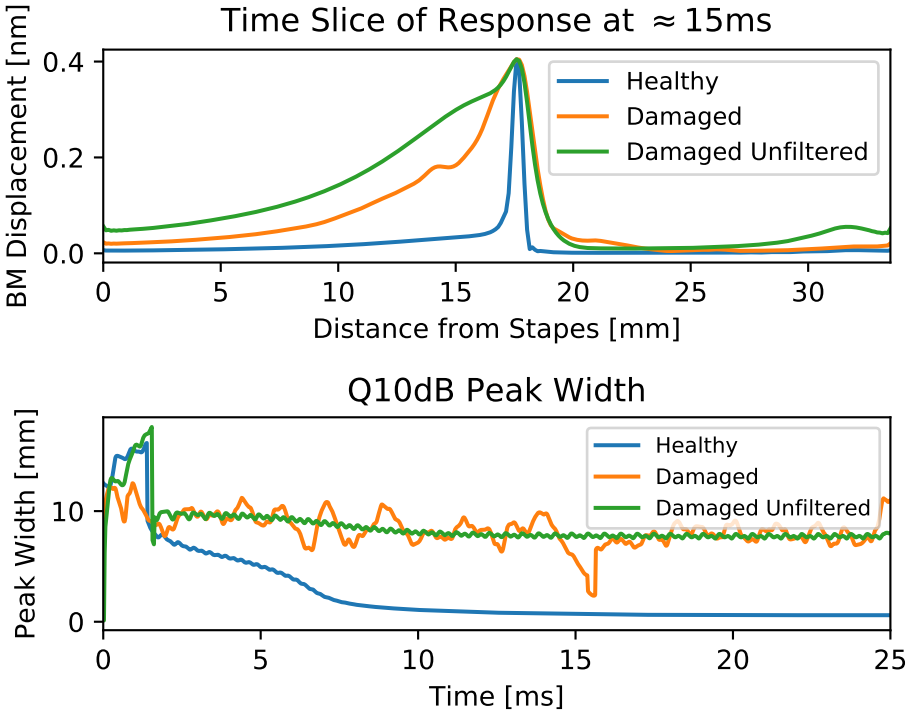

Figure 4.14: Optimization result for 60% uniform hearing damage.

REFERENCES

[1] J. S. Oghalai, "The cochlear amplifier: augmentation of the traveling wave within the inner ear," *Current opinion in otolaryngology & head and neck surgery*, vol. 12, pp. 431–438, Oct. 2004.

[2] F. R. Lin, "Hearing loss prevalence in the United States," *Archives of Internal Medicine*, vol. 171, pp. 1851–1852, Nov. 2011.

[3] C. Yoshinaga-Itano and M.-r. L. Apuzzo, "Identification of hearing loss after age 18 months is not early enough," *American Annals of the Deaf*, vol. 143, no. 5, pp. 380–387, 2012.

[4] S. Arlinger, "Negative consequences of uncorrected hearing loss–a review," *International Journal of Audiology*, vol. 42, no. Supplement 2, pp. 17–20, 2003.

[5] C. M. Li, X. Zhang, H. J. Hoffman, M. F. Cotch, C. L. Themann, and M. R. Wilson, "Hearing impairment associated with depression in US adults, National Health and Nutrition Examination Survey 2005-2010," *JAMA Otolaryngology - Head and Neck Surgery*, vol. 140, no. 4, pp. 293–302, 2014.

[6] H. Dillon, *Hearing Aids*. Hodder Arnold, 2008.

[7] B. Edwards, "The future of hearing aid technology," *Trends in Amplification*, vol. 11, no. 1, pp. 31–45, 2007.

[8] M. C. Liberman, J. Gao, D. Z. Z. He, X. Wu, S. Jia, and J. Zuo, "Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier," *Nature*, vol. 419, pp. 300–304, Sep. 2002.

[9] M. Rudnicki, O. Schoppe, M. Isik, F. Völk, and W. Hemmert, "Modeling auditory coding: from sound to spikes," *Cell and Tissue Research*, vol. 361, pp. 159–175, Jul. 2015.

[10] W. C. Stebbins, J. E. Hawkins, L. G. Johnson, and D. B. Moody, "Hearing thresholds with outer and inner hair cell loss," *American journal of otolaryngology*, vol. 1, no. 1, pp. 15–27, 1979.

[11] B. C. Moore, "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," *Ear and hearing*, vol. 17, pp. 133–161, Apr. 1996.

[12] M. Florentine, S. Buus, B. Scharf, and E. Zwicker, "Frequency selectivity in normally-hearing and hearing-impaired observers," *Journal of Speech, Language, and Hearing Research*, vol. 23, pp. 646–669, Sep. 1980.

[13] M. Müller and J. W. T. Smolders, "Shift in the cochlear place-frequency map after noise damage in the mouse," *Neuroreport*, vol. 16, pp. 1183–1187, Aug. 2005.

[14] H. Von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Longmans, Green, 1912.

[15] R. Nobili and F. Mammano, "Biophysics of the cochlea II: stationary nonlinear phenomenology," *The Journal of the Acoustical Society of America*, vol. 99, pp. 2244–2255, Apr. 1996.

[16] E. G. Wever, C. W. Bray, and M. Lawrence, "The interference of tones in the cochlea," *The Journal of the Acoustical Society of America*, vol. 12, pp. 268–280, Oct. 1940.

[17] L. Grafakos, *Classical Fourier Analysis*, vol. 2. Springer, 2008.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

[19] R. G. Regis and C. A. Shoemaker, "A stochastic radial basis function method for the global optimization of expensive functions," *INFORMS Journal on Computing*, vol. 19, pp. 497–509, Nov. 2007.

[20] A. Costa and G. Nannicini, "RBFOpt: an open-source library for black-box optimization with costly function evaluations," *Mathematical Programming Computation*, vol. 10, pp. 597–629, Dec. 2018.

[21] H.-M. Gutmann, "A radial basis function method for global optimization," *Journal of global optimization*, vol. 19, no. 3, pp. 201–227, 2001.

[22] M. Florentine, H. Fastl, and S. Buus, "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 195–203, 1988.

[23] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: a generative model for raw audio," *arXiv preprint arXiv:1609.03499*, pp. 1–15, 2016.

[24] W. Kim, S. Kim, S. Huang, J. S. Oghalai, and B. E. Applegate, "Picometer scale vibrometry in the human middle ear using a surgical microscope based optical coherence tomography and vibrometry system," *Biomedical Optics Express*, vol. 10, pp. 4395–4410, Sep. 2019.

APPENDIX A

UNITS

A.1 Converting dB SPL to stapes acceleration

In the context of hearing-related study, sound intensity is often measured in therms of decibels sound pressure level or $\mathrm{dB\ SPL}$. However, the cochlea model in §2.2 accepts the input in terms of stapes acceleration. Mammano & Nobili give a reference point that $5\,\mathrm{m/s^2}$ stapes acceleration is roughly equivalent to $80\,\mathrm{dB\ SPL}$ in the $1\,\mathrm{kHz}$ to $2\,\mathrm{kHz}$ range[15]. The tympanic membrane displacement as a function of sound intensity in $\mathrm{dB\ SPL}$ is known[24]. If we assume that the displacement of the stapes is on the order of the displacement of the tympanic membrane, we can approximate the magnitude of stapes acceleration as a function of $\mathrm{dB\ SPL}$ and vice verse.

The tympanic membrane displacement (d) is linear with sound intensity in $\mathrm{dB\ SPL}$, and it is safe to assume that the stapes acceleration (a) is also linear with input intensity:

$$SPL_{in} = m_1 d + b_1 \tag{A.1}$$

$$SPL_{in} = m_2 a + b_2 \tag{A.2}$$

So

$$m_1 d + b_1 = m_2 a + b_2 \tag{A.3}$$

When tympanic membrane displacement (d) is zero, the stapes acceleration (a) must also be zero, so we know that $b_1 = b_2 = b$. From fitting a line to the data in [24], we find that $m_1 = 236.67$, $b = 1.1246$, and $d = 42.2482$ at $80\,\mathrm{dB\ SPL}$. Keeping in mind that $a = 5\,\mathrm{m/s^2}$ at $80\,\mathrm{dB\ SPL}$, we can use these values to find m_2:

$$m_2 = \frac{236.67 * 42.2482}{5} \approx 2000 \tag{A.4}$$

Remembering that decibels are a logarithmic unit, we need to use the following relation:

$$SPL_{in} = 10^{S/20} \tag{A.5}$$

The final relationship for converting stapes acceleration in $\mathrm{m/s^2}$ to $\mathrm{dB\ SPL}$ is then given in Equation A.6 where a is the stapes acceleration and S is the sound amplitude in $\mathrm{dB\ SPL}$.

$$S = 20 \log_{10}(2000a + 1.1246) \tag{A.6}$$

We can also solve this equation for a in terms of S so that we may convert from $\mathrm{dB\ SPL}$ to $\mathrm{m/s^2}$:

$$a = \frac{10^{S/20} - 1.1246}{2000} \tag{A.7}$$

A.2 Input Conversion Factor

The computation in the Mammano & Nobili cochlea model described in §2.2 employs some non-standard units. They define the standard unit of distance as their own β unit, where one β is approximately equal to the length of the human cochlea which they assume to be $33.5\,\mathrm{mm}$. The time unit in the simulation is assumed to be ms. So, the simulation assumes units of β, kg, and ms.

If the stapes acceleration in β/ms is a_s and dT is the sample interval in ms, then the model input is taken to be:

$$I = a_s \times dT \tag{A.8}$$

The final input is in units of β/ms. To convert the input acceleration from $\mathrm{m/s^2}$ to β/ms, a scaling factor

$$I[\beta/\mathrm{ms}] = a_s[\mathrm{m/s^2}] \times \frac{dT[\mathrm{ms}]}{33.5\,\mathrm{mm}} \tag{A.9}$$

Conventionally, the sample rate is $44.1\,\mathrm{kHz}$, so $dT = 1/44.1\,\mathrm{kHz} \approx 0.0227\,\mathrm{ms}$. So, the final scaling factor is:

$$F = \frac{1/44.1}{33.5e3}\,[\mathrm{ms}/\beta] \approx 6.769 \times 10^{-7}\,[\mathrm{ms}/\beta] \tag{A.10}$$

The final input signal should be calculated as follows before running the model:

$$I = a_s \times F \tag{A.11}$$