Analysis of spectro-temporal receptive fields in an auditory neural network

Madhav M. Nandipati

Abstract

Neural networks have been utilized in a wide variety of fields. Extensive research relating to the application of neural networks in computational biology has been done, but the validity of these biological models has yet to be established. Models of the auditory cortex serve as examples of this problem; in these models, the properties of artificial neuronal populations are hard to fully characterize with traditional methods such as tuning curves. Spectro-temporal receptive fields (STRFs) are able to characterize neurons in both the spectral and temporal domains, giving them greater power to analyze the properties of neural networks. STRFs have been obtained in a variety of animals, but have not been adequately studied in computational models. The aim of this project is to generate STRFs on a basic neural network model of the early auditory processing stages in the brain. The characteristics of the STRFs obtained from this present neural network are remarkably similar to those from the biological system and reveal some of the processing that takes place in the auditory cortex.

Keywords: neural network, auditory processing, spectro-temporal receptive field (STRF)

1 Introduction

Recent computational models have employed a neural network architecture to mimic the processing that occurs in the brain. Although many of these models revealed how well we understand the brain, verifying the biological realism remains difficult. In this project, I discuss the use of receptive fields in a neural network and application of receptive fields to assess the validity of computational models.

The brain is a remarkably complex organ that is responsible for the intricate processing of tactile and abstract information. The hominine capability of sight has already been widely studied and well documented in research and modeling. Yet our knowledge of auditory processing in the brain is surprisingly limited. Realistic models of the auditory cortex will give researchers the tools to better understand and mimic the brain's diverse functions in artificial systems.

Spectro-temporal receptive fields (STRFs) are visual descriptions of the linear properties of auditory neuronal populations. STRFs accurately describe both the spectral (frequency) and temporal (time) components of neuronal responses. With receptive fields, computational models can be studied in greater detail. The computer-generated STRFs are hypothesized to be able to evaluate the realism of auditory processing models of the brain. To this effect, I have employed the use of a newly developed, neural network model of the brain as the preferred way to model early auditory processing stages and generate STRFs. The model uses a simple representation of memory to approximate temporal processing in the auditory cortex. The connection factors, or weights, between different neurons determine how the neurons respond to auditory stimuli. These weights were fashioned through an unsupervised training algorithm using a training set composed of frequency-modulated (FM) sweeps and pure tones. Using this training procedure, the neural network extracted some of the statistical regularities from the training set . Subsequently, complex, moving ripple stimuli were used in the model to obtain the receptive fields of the neuronal population.

The resulting receptive fields illustrate the properties of the neural network. By analyzing STRFs, the validity of the computational model can be determined. These receptive fields can also be compared against the tuning curves to investigate the quality of the information each method illustrates.

As researchers strive to develop ever more realistic, computational models of the brain, the detail and elegance of receptive fields will equip researchers

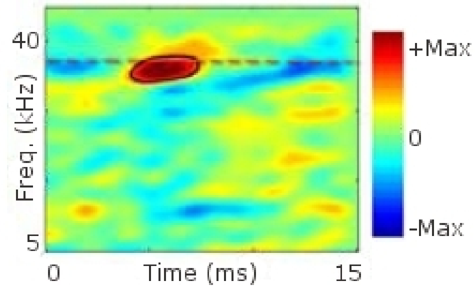

Figure 1: Biological STRF from a Mexican free-tailed bat with inhibition blocked. The red represents the area with maximum neuronal activity. See References for source of image[?]

to better evaluate the models against their biological counterparts.

2 Background

2.1 Layout of the ear

The ear is the earliest stage of auditory processing. The ear is divided into three main areas: the outer ear, the middle ear, and the inner ear. The general purpose of the outer and middle ear is to convey and amplify the mechanical vibrations in the air (sound) to the inner ear. Transduction, the process of converting mechanical signals into electrical potentials, takes place in the inner ear[?] . The vibrations in the inner ear selectively cause hair cells along the basilar membrane in the cochlea to move[?] . The motion of the hair cells allows electrical potentials to travel to the auditory nerve and become processed by the brain. Hair cells are theorized to be frequency-selective. Specific pitches excite specific areas of the basilar membrane, information which is relayed to the cortical levels of auditory processing[?] .

2.2 Oja's rule

Unsupervised learning paradigms allow neural network models to dynamically modify their own weighted connections between nodes, analogous to the changes in synaptic plasticity between neurons [?] . The simplest form of unsupervised training is based on the Hebbian learning rule[?] . Hebb hypothesized that if two neurons are simultaneously active, the connection between them would be strengthened [?] . As a mathematical equation, Hebb's rule can be represented as:

$$\Delta_t w_{ij} = \epsilon x_i y_j \tag{1}$$

where $\Delta_t w_{ij}$ represents weight change between two units, ϵ is the learning rate, and x_i and y_j are the activation values of the pre-synaptic and post-synaptic neurons, respectively[?] . Hebb's rule is a concise, albeit very limited, simplification of the synaptic plasticity of neurons.

Hebb's rule is inherently unstable. One overwhelming problem with Hebb's rule is that the weights diverge to infinity after repeated iterations[?] . Oja's rule, a modified version of Hebb's rule, fixes this problem by subtracting a portion of the existing weight away from the weight change[?] . Oja's rule can be shown as:

$$\Delta_t w_{ij} = \epsilon(x_i y_j - y_j^2 w_{ij}) \tag{2}$$

where $\Delta_t w_{ij}$ represents weight change between two units, w_{ij} is the current weight, ϵ is the learning rate, and x_i and y_j are the activation values of the pre-synaptic and post-synaptic neurons, respectively[?][?] . The learning rate is a key parameter that dictates how quickly the weights are updated. While a very small learning rate will cause the weights to change slowly over the training set, a sufficiently large learning rate will cause the weights to oscillate. In this project, the learning rate was first set at 0.015 and then further decreased in order to avoid these two issues.

2.3 Spectro-temporal receptive fields (STRFs)

STRFs represent the linear properties of primary auditory processing neurons and depict the neuronal impulse response characterizations at frequency-time

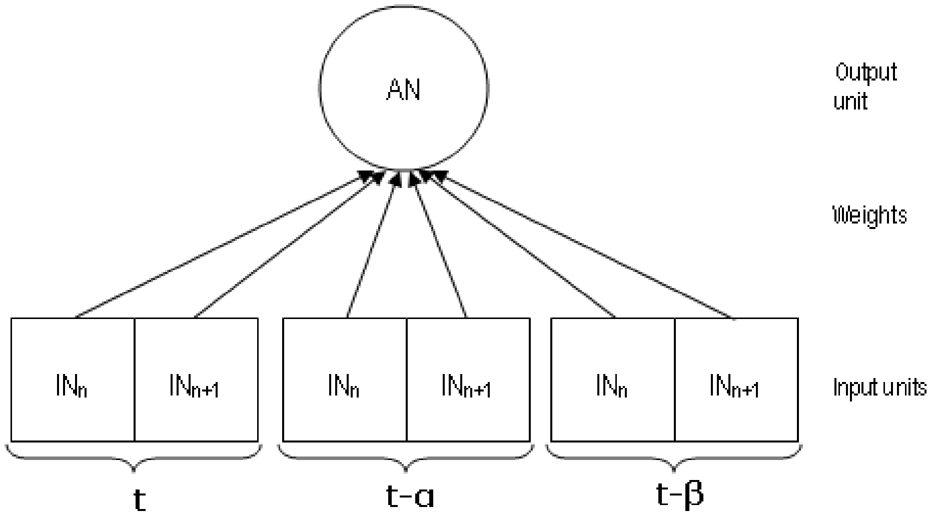

Figure 2: Single artificial neuron of neural network. Each artificial neuron in the second layer is connected to three timesteps: the current timestep (t), timestep from 24 ms ago ($t - \alpha$), and timestep from 48 ms ago ($t - \beta$). The weights between these timesteps and the artificial neuron were trained with Oja's Rule.

points[? ? ? ? ? ? ?] . STRFs are generated by collecting a neuron's responses to different moving ripple stimuli. Since these stimuli are approximate components of complex sounds, the STRFs characterize the neuron response to spectro-temporally rich sound stimuli[? ?] . Figure 1 shows an image of a biological STRF from Mexican free-tailed bats.

Since STRFs describe the neuronal responses in both the spectral and temporal dimensions, they are hypothesized to be more useful than traditional methods of describing neurons such as tuning curves. Tuning curves only depict a neuron's spectral property. STRFs have been used to predict the outputs of neurons, further illustrating the utility of receptive fields in the auditory world[? ? ?] .

3 Methods

3.1 Neural network

Neural networks mimic biological processing by joining layers of artificial neurons in a meaningful way. While real neurons rely on neurotransmitter communication, the artificial neurons in neural networks transfer information through weighted connections. The neural network employed in this project is a two-layer model that responds to sound stimuli. The input to the model is a spectrogram, frequency vs. time distribution of sound, in order to account for frequency-selectivity in the basilar membrane. The spectrogram is split into distinct time and frequency intervals, where each time interval represents 12 ms of auditory information and each frequency interval represents a 43 Hz range of the auditory stimulus. Each distinct time interval is known as a timestep.

This model utilizes time-delayed inputs to respond to different timesteps of the auditory stimulus. The first layer of the neural network contains the information from a timestep (t), the previous timestep from 12 ms ago ($t - \alpha$), and the previous timestep from 24 ms ago ($t - \beta$). Each of these timesteps composes the entire first layer, which becomes 387 units long (three timesteps each of 129 units).

The second layer of the neural network is the output layer. The auditory information gets relayed to the second layer through a network of connection factors. These connections are weighted so that some parts of the input contribute to the overall output more than other parts. The value of a second layer artificial neuron can be summarized in the following equation:

$$y_j = \sum_{i=1}^{129} w_{ij} x_i + \sum_{i=1}^{129} w_{(i-\alpha)j} x_{i-\alpha} + \sum_{i=1}^{129} w_{(i-\beta)j} x_{i-\beta} \tag{3}$$

where w is the weight matrix, x_i is the input from the current timestep, $x_{i-\alpha}$ is the input from the timestep that appeared 12 ms ago, and $x_{i-\beta}$ is the input from the timestep that appeared 24 ms ago.

Using this configuration, the neural network was trained using Oja's rule. The weights between the first and second layers of the model were originally set at zero-centered, normally distributed, random values. Though the weight values become modified by training using Oja's Rule, the connections between the artificial neurons are neither created nor eliminated. Each artificial neuron in the second layer is permanently connected to a set of input units.

Animals can only hear a very limited range of frequencies. In this project, neural network artificially simulates a limited range of hearing to more accurately match the real world. The model is restricted to frequencies between 172 and 5512 Hz by removing any weighted connections between the second layer and the first four frequency bins of a timestep. See Figure 2 for a schematic of the network.

3.2 Unsupervised training

The neural network was trained on a simple set of frequency-modulated (FM) sweeps and pure tones. Each training stimulus consisted of either an upsweep, pure tones, or downsweep. Each upsweep had an initial frequency between 0.17 and 2.250 kHz. Each pure tone had an initial frequency between 2.27 and 3.43 kHz. Each downsweep had an initial frequency between 3.45 and 5.51 kHz. This training set was sampled at a rate of 11.025 kHz. According to the Nyquist-Shannon Sampling Theorem, a sampling rate of 11025 Hz reduces the maximum possible frequency to 5.512 kHz[?] .

Afterwards, these stimuli were converted into spectrogram matrices (frequency vs. time) and scaled between 0 and 1 in Matlab[?] . These matrices served as the input training data to the neural network. The training set was presented to the neural network in a random order. After each presentation of a single timestep, the model modified its own weights according to Oja's rule. Only one or two artificial neurons were trained at one time because most of the frequency-time space of the spectrogram is empty, and the model would be mostly trained on an empty stimulus. The initial frequency of the training stimulus determined which artificial neurons were trained. After 50 iterations through the training set, the learning rate was decreased by a factor of 2 to prevent the weights from oscillating around a value. The learning rate was decreased 12 times, so the model iterated through the training set of 275 spectrograms a total of 600 times. The weights were saved for further use.

3.3 Constructing stimuli

3.3.1 Moving ripples

The moving ripple stimuli are complex, broadband noises that are used to determine the spectro-temporal receptive fields (STRFs) of artificial neuronal populations[? ? ? ? ? ? ?] . The moving ripples are composed of hundreds of densely packed pure tones log-spaced between 0.975 to 5.512 kHz that were sinusoidally modulated in both the spectral and temporal domains. The ripple equation, intensity at specific frequency-time points, is given as:

$$S(t,x) = 1 + \Delta A \times sin[2\pi(\omega t + \Omega x) + \Phi] \qquad (4)$$

where $S(t,x)$ is intensity, t is time, $x = log_2(F/F_0)$ where x is the logarithmic frequency axis, F_0 is the baseline frequency, F is the frequency, ΔA is modulation depth, ω is the ripple velocity (Hz), Ω is the ripple frequency (cycles/octave), and Φ is the phase shift (radians)[? ? ? ? ?] . The stimuli were generated using a Matlab script[?] .

These ripple stimuli were varied across two parameters separately, the ripple velocity (Hz) and the ripple frequency (cycles/octave). The ripple velocity was varied from -40 to 0 Hz in steps of 2 Hz and the ripple frequency was varied from 0.0 to 4.0 in steps of 0.4 cycles/octave. In total, 231 ripple stimuli were used to obtain the spectro-temporal receptive fields. See Figure 3 for four representative spectrograms of the ripple stimuli set. The other parameters were held constant for all ripple stimuli.

The output of the neural network units to the different moving ripples were computed. The transfer function (TF) can be created from the artificial neuronal responses. The TF is a broad characterization of a unit's responses to the ripple stimuli and is defined by:

$$TF(\omega,\Omega) = M(\omega,\Omega) \times exp[i \times \Phi(\omega,\Omega)] \qquad (5)$$

where $i = \sqrt{-1}$, $\Phi(\omega,\Omega)$ is the response phase (radians), and $M(\omega,\Omega)$ is the response magnitude[?] . In order to construct the TF, the magnitude and phase of the raster responses were calculated by performing a Fourier transform. The magnitude was the maximum value of the first half of the transform, and

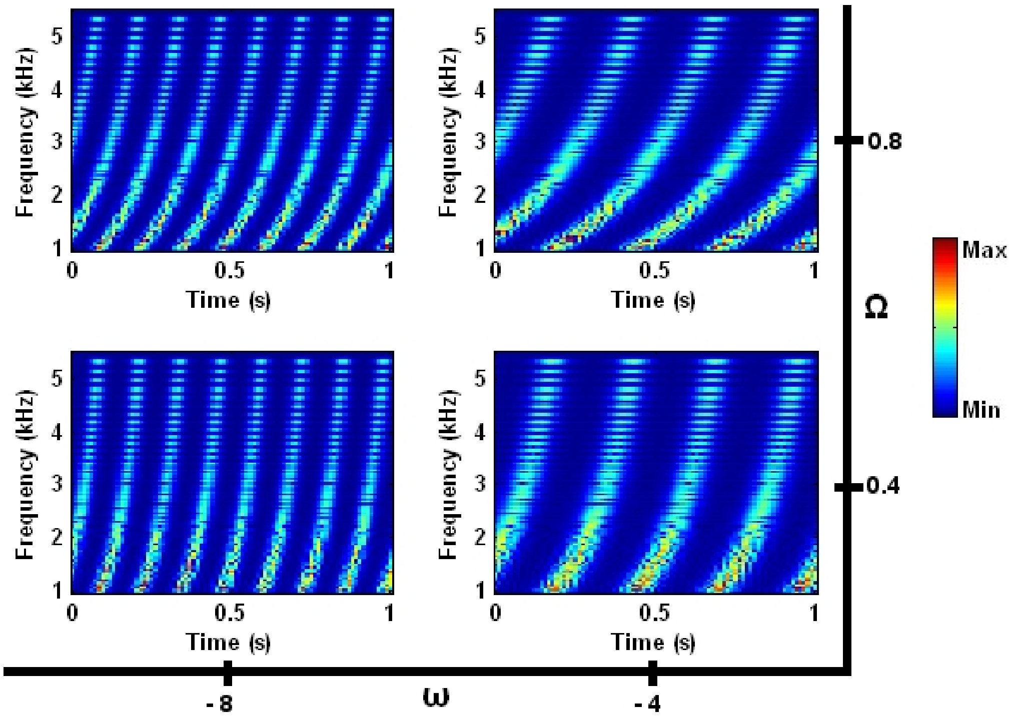

Figure 3: Four ripple stimuli. The ripples in this figure are varied across ω with Ω held constant at 1.0 cycles/octave. Positive ω corresponds to downward moving ripples. These stimuli were used to generate the receptive fields of the artificial neurons.

the phase was extracted from the unwrapped angle at that point. The first two values of the Fourier transform were discarded because each neuronal response contained the low frequencies represented by the first two values of the transform. Also, the second half of the Fourier transform was discarded because it provides redundant information.

A two-dimensional inverse Fourier transform function was performed on the transfer function in order to generate the desired STRF[?, ?, ?, ?]. Before the transformation, the transfer function was padded with zeros to smooth the resulting receptive field.

3.3.2 Tuning curve tones

Tuning curves have been used extensively in both biological and computational applications because they allow researchers to quantitatively analyze the frequencies at which a specific auditory neuron responds best to. To generate these curves, the firing rates of the neurons are collected in response to pure tones varied across the frequency domain[?]. The neurons respond with the greatest intensity to tones that match their best frequency (BF) and with decreasing intensity to tones away from their BF. The plots of these rates against the frequency of tone generate the tuning curves.

In this project, the tones that were used to construct the tuning curves were generated in Matlab. These tones were 1 second long, and sampled at identical settings to the training set, and were subsequently converted to spectrograms to become the input of the neural network. The frequencies of the tones were varied from 10 to 5490 Hz in steps of 40 Hz. The responses of the artificial neurons to these tones were collected. The maximum response to each tone was plotted in a intensity vs. frequency plot, and the peak of the plotted curve denotes the BF of the artificial neuron.

4 Results

4.1 Receptive fields

The receptive fields for the six different artificial neurons are plotted in Figure 4. The abscissa represents time after stimulus onset and the ordinate represents

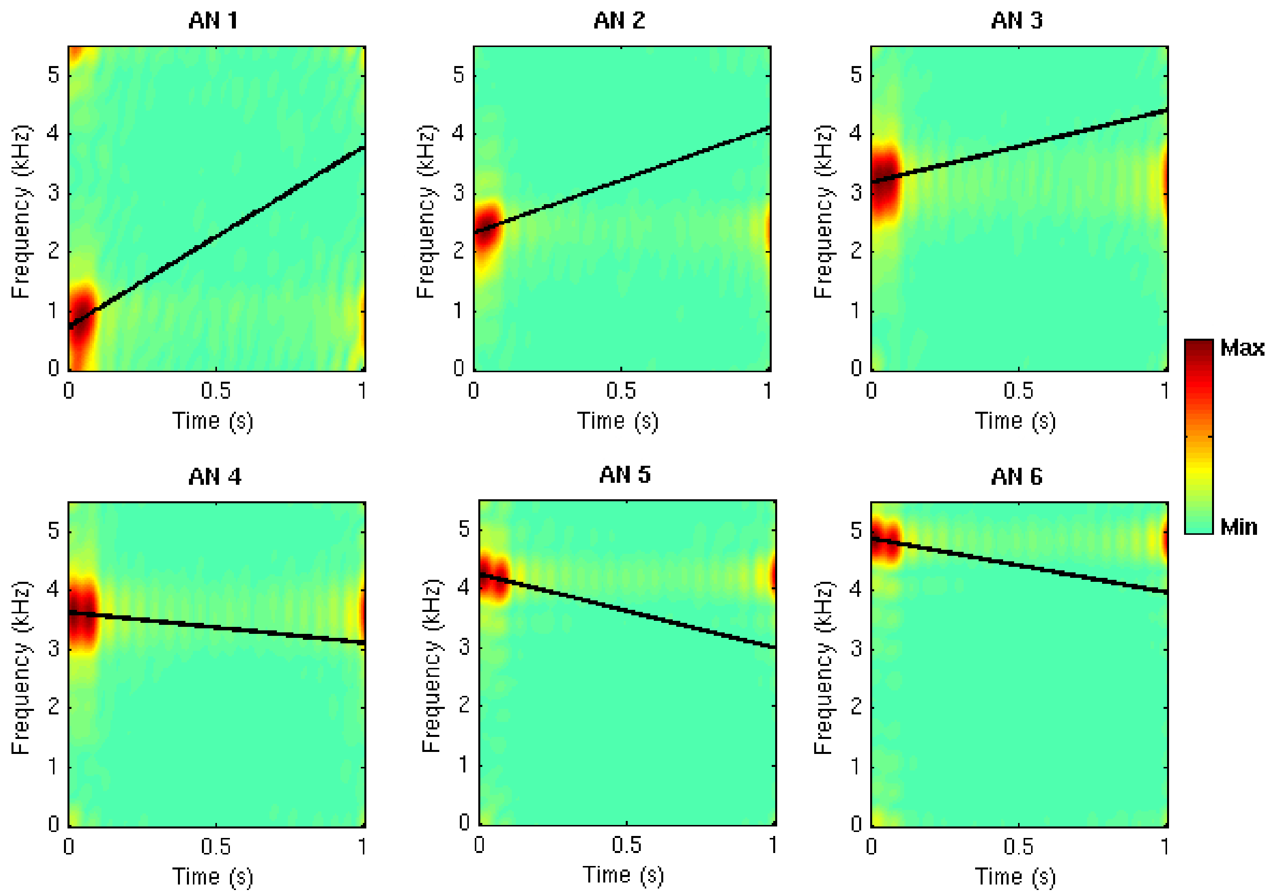

Figure 4: Spectro-temporal receptive fields (STRFS) from neural network. The STRFs from the artificial neurons (AN) demonstrate frequency selectivity. The width of the area of activation depicts the temporal properties of the artificial neuron. The black line illustrates the angle of the receptive field.

frequency. The red areas of the graph shows where the artificial neuron responds with greatest intensity. The blue areas shows where the artificial neuron does not respond to, or responds very weakly to. The graphs show that the artificial neurons are responding to distinct frequency ranges. For instance, the STRF for AN 1 shows that the second artificial neuron responds strongly to a frequency of 877 Hz. As the artificial neurons are connected to higher frequency input units, the receptive fields show that the artificial neuron also responds to higher frequencies. This result agrees with the hypothesized outcome.

To quantitatively analyze the frequency properties of the artificial neuron, the best frequency (BF) can be obtained from the STRFs. The spectral component of the maximum value of the STRF represents the frequency at which an artificial neuron responds best. The BFs for all artificial neurons are shown in Table 1.

The width of the receptive field shows how long the neuron responds to a complex stimuli. The STRFs show that the neural network responds strongly to a givens stimulus over a short time period. Since the artificial neurons were trained on different types of stimuli, the STRFs reflect the varying properties that arose from the training. The black line over the STRF denotes the general angle of the receptive field. This angle was determined by performing a linear regression on the significant areas of the STRF. The pixels considered significant were greater than 7 standard deviations from the mean. The inverse tangent of the slope of the best fit line gives the angle of the STRF.

The angle for each artificial neuron is given in Table

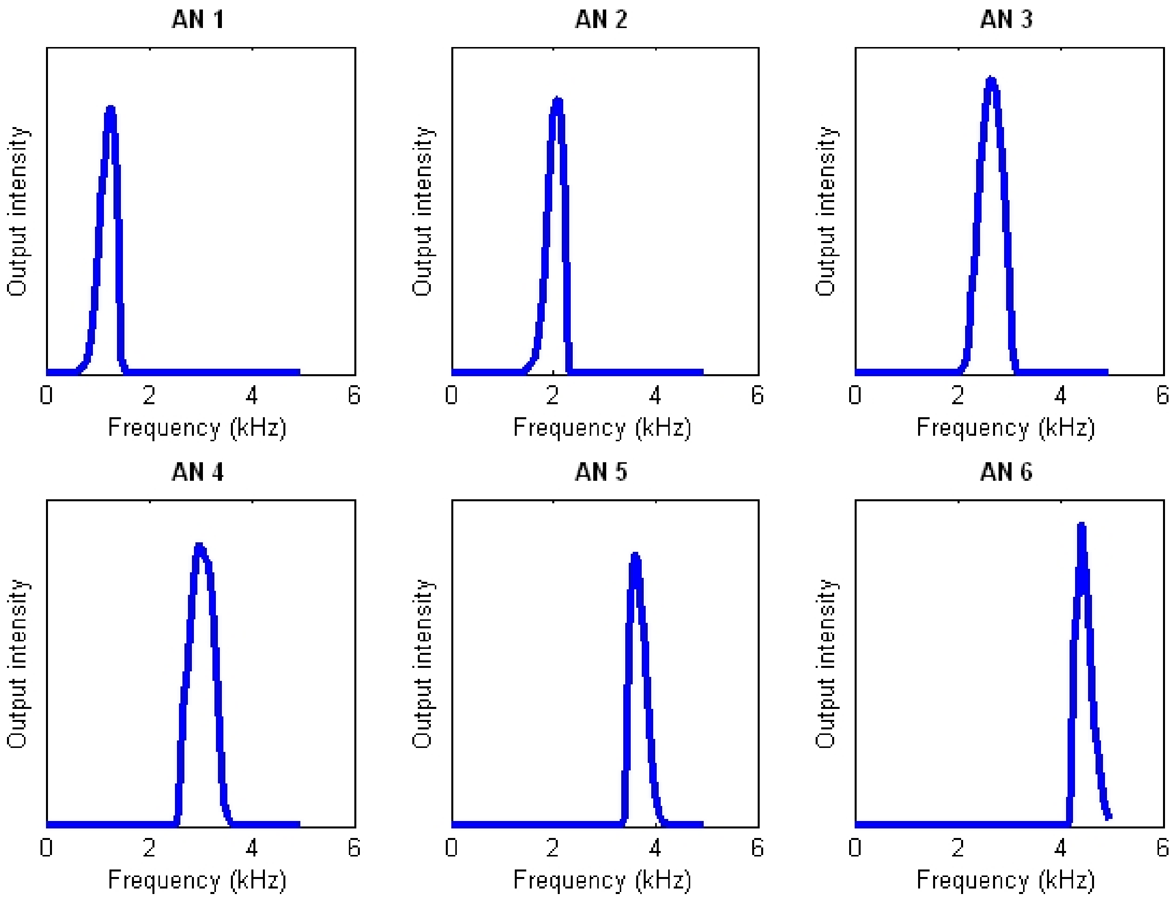

Figure 5: Tuning curves from neural network. The artificial neurons (ANs) are selective for frequencies where response intensity is greater than zero. The peak of the curves gives the best frequency (BF) of the AN. These graphs do not describe the artificial neurons in the temporal domain.

2. For instance, the first two neurons were trained on a set of upsweeps and the STRF for each of those artificial neurons exemplifies the nature of that training set; the third neuron was trained on a combination of upsweeps and pure tones, so the angle of the STRF is less than the angle from the first two STRFs. The existence of an angled receptive field further illustrates the temporal properties of the neural network.

4.1.1 Tuning curves

After collecting the responses of the artificial neurons to the pure tones, the tuning curves were obtained. The graphs of the tuning curves in intensity vs. frequency of all the artificial neurons are shown in Figure 5. Although biological neurons can only respond at one fixed strength (all-or-none principle), the intensity of the response can be quantified as the firing rate, or how many times a neuron fires per time unit. The intensity of the response in the neural network is analogous to the firing rate in biological neurons. The maximum of the tuning curve is BF of the neuron[?] .

The tuning curves, similar to the receptive fields, show that the artificial neurons respond to specific frequency ranges. The BFs from the STRFs is closely correlated (r=0.9818) with the BFs from the tuning curves. The tuning curves, though, do not give any indication of how the artificial neurons are responding over time.

5 Discussion

Receptive fields have been used in this project to establish the linear properties of neural networks. Another goal of the project was to compare results from

Table 1: BFs for the artificial neurons.

AN	AN 1	AN 2	AN 3	AN 4	AN 5	AN 6
BF (kHz)	0.877	2.406	3.257	3.639	4.231	4.866

Table 2: Angles for the artificial neurons' receptive fields.

AN	AN 1	AN 2	AN 3	AN 4	AN 5	AN 6
Angle (deg)	29.09	18.00	12.55	-5.39	-12.86	-9.43

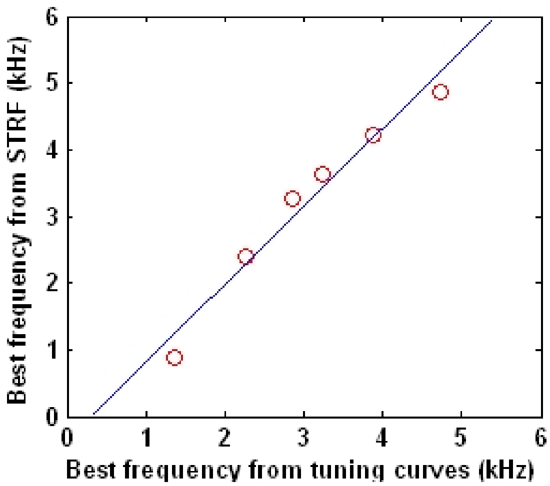

Figure 6: Comparison between the BFs from the STRFs and the tuning curves.

the STRFs and tuning curves. The high correlation of BFs found using STRFs and tuning curves validates the spectral components of STRFs. More importantly, the STRFs also characterize the temporal properties of the artificial neuronal population, shown by the different angles of the artificial neurons. Tuning curves, on the other hand, can only depict the spectral properties. The STRFs allow researchers to validate a neural model because the STRFs are able to show the linear properties of the neurons, unlike traditional methods of characterizing neural networks. The close resemblance between the artificial and biological STRFs hints at the type of processing done the auditory cortex. Early auditory processing stages must also process sound stimuli with neurons responding to different frequencies. As the neural network shows, neuronal responses are functions of previous inputs to the model to perform temporal processing.

5.1 Limitations

This neural network was constructed to evaluate STRFs and compare them to tuning curves in computational models, and so it is only a basic model. Continued work will need to be done to improve the realism of the model. The neural network does not attempt to accurately simulate the spiking rate, so the artificial neuronal response intensity is only compared relative to other neurons. Lateral inhibition, a phenomenon where neurons inhibit the response of other neurons, was not modeled in the present neural network. Biological STRFs depict varying levels of lateral inhibition. Further additions to the model may produce the lateral inhibition patterns that are found in biological STRFs.

6 Conclusion

STRFs can be utilized to analyze the properties of computational models of auditory processing in a visual manner. The receptive fields describe how an artificial neuron would generally respond to both the spectral and temporal aspects of sound stimuli. This characterization of the neural network can then assist researchers in determining the accuracy of their models, without reading lines of code or examining multiple outputs from each individual artificial neuron. The STRFs from the auditory network would be able to show if a model is responding to the correctly to time-delayed inputs and frequency ranges in just a few graphs. Tuning curves are also able to depict the realism of models, but not with respect to both frequency and time information. In this way, the simpler tuning curves do not have the same ability of receptive fields.

Neural networks and receptive fields also have potential to help researchers and physicians. The most prevalent auditory disorder is hearing loss, and neural networks with receptive fields may be able to aid physicians in characterizing hearing loss and its causes. With further development, this model can approximate some aspects of auditory processing, specifically relating to processing in the cortex, and can be used to simulate hearing loss and other diseases associated with it such as tinnitus. Tinnitus is characterized by a ringing in the ears in the absence of external stimuli. Researchers would be able to retrain the weights of the model using Oja's Rule by using stimuli such as loud music and other destructive noises, and then generate the STRFs with the new weights. If the weighted connection becomes greater than an arbitrary threshold, then the weight is set to zero, modeling the loss to hear certain frequencies. By examining the receptive fields, researchers can tell which parts of the brain are affected by the sounds and how both the spectral and temporal properties change because of the hearing loss. This neural network may be able to track the effect of high frequency hearing loss on the normally heard low frequencies. Researchers would explore hypotheses relating to the effect of the potentially dangerous effects of sounds without actually exposing humans.

The neural network could also serve as a starting point in investigating new therapies for hearing loss. One possible cortical level treatment for hearing loss is electrode stimulation, where small electrical currents are delivered to parts of the brain. Another type of treatment is transcranial magnetic stimulation (TMS). Both of these therapies can be first tested in neural networks by simulating the effect of brain stimulation. Researchers would be able to quickly validate the use of the therapy by examining the properties of the artificial neurons through STRFs.

The use of spectro-temporal receptive fields in neural networks is now emerging. This project demonstrates the flexibility and utility of STRFs in visually describing the properties of artificial neurons in an auditory model and determining the validity of computational models. In the future, this research may be extended to studying hearing loss and other auditory disorders. The insights gained from models may be used to design better focused experiments and evaluate novel therapies.

7 Acknowledgements

This work has only been possible with the guidance and advice of my mentors at the Brain and Imaging Modeling Section of the National Institute on Deafness and Other Communication Disorders, of the National Institutes of Health. I sincerely thank Dr. Barry Horwitz, Dr. Fatima Husain, and Dr. Jason Smith. I also thank Kristina Visscher of Brandeis University for providing the script to generate the ripple stimuli. I am grateful for the comments and suggestions provided by my teacher, Mr. Randolph Latimer.

References

[] Satoshi Amagai, Robert J. Dooling, Shihab Shamma, Tracy L. Kidd, and Bernard Lohr. Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars (Melopsittacus undulatus). *Journal of the Acoustical Society of America*, 105(3):2029–2035, 1999.

[] Sari Andoni, Na Li, and George D. Pollak. Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *Journal of Neuroscience*, 27(18):4882–4893, 2007.

[] Didier A. Depireux, Jonathan Z. Simon, and David J. Kleinand Shihab A. Shamma. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of Neurophysiology*, 85:1220–1234, 2001.

[] D. O. Hebb. *The Organization of Behavior*. Lawrence Erlbaum, 2002.

[] Christoph Kayser, Christopher I. Petkov, and Nikos K. Logothetis. Tuning to sound frequency in auditory field potentials. *Journal of Neurophysiology*, 98(3):1806–1809, 2007.

[] Nina Kowalski, Didier A. Depireux, and Shihab A. Shamma. Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra. *Journal of Neurophysiology*, 76:3503–3523, 1996.

[] Nina Kowalski, Didier A. Depireux, and Shihab A. Shamma. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology*, 76(5):3503–3523, 1996.

[] MathWorks. Matlab, 2007.

[] Randall C. O'Reilly and Yuko Munakata. *Computational Explorations in Cognitive Neuroscience*. MIT Press, 2000.

[] Dale Purves, George J. Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara, and S. Mark Williams. *Neuroscience*. Sinauer Associates, 2004.

[] Powen Ru. Ripple, 1998.

[] Sebastian Seung. Unsupervised learning, 2002. Lecture.

[] Shihab Shamma. On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5(8):340–348, 2001.

[] Claude E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[] Kristina M. Visscher, Elina Kaplan, Michael J. Kahana, and Robert Sekuler. Auditory short-term memory behaves like visual short-term memory. *PLoS Biology*, 5(3):662–672, 2007.