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

Madhav Nandipati

Abstract

Neural networks have been utilized for a vast range of applications, including computational biology. But the realism of these models remains in question. In the models of the auditory cortex, for example, the properties of neuronal populations are hard to fully characterize with traditional methods such as tuning curves. Spectro-temporal receptive fields (STRFs), which describe neurons in both the spectral and temporal domains, have been obtained in a variety of animals, but have not been adequately studied in computational models. The aim of this project is to address these issues by generating the spectro-temporal receptive fields on a basic, neural network model of the early auditory processing stages in the brain. This novel use of STRFs is also evaluated against tuning curves in describing the properties of the neural network.

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

1 Introduction

Neural networks are utilized for a wide range of applications in both mathematics and science, and recent efforts have been made to mimic the processing that occurs in the brain. Although many of these models revealed how well we understand the brain, the realism of such models remains in question. In this paper, 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 the intrigue of sound and the auditory processing in the brain remain is still surprisingly new territory. Realistic models of the auditory cortex will give scientists the tools to better understand and hopefully mimic these complex 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 two newly developed, neural network models of the brain as the preferred way to model early auditory processing stages and generate STRFs. One model is uses a linear transform scheme and the other model has a simple representation of memory. 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 natural stimuli. Using this training procedure, the neural network extracted some of the statistical regularities of the natural world. Subsequently, complex, moving ripple stimuli were used in the model to obtain the receptive fields of the neuronal populations.

The resulting receptive fields illustrate the properties of the neural network. By analyzing the properties of the model, the validity of the neural network can be determined. To investigate this, other neural networks were created using the untrained and trained weights. The different weight matrices serve as different levels of realism that can easily be implemented in the same model. The comparison of the receptive fields from the two different models can be used to evaluate the validity of the neural network.

As scientists strive to develop ever more realistic, computational models of the brain, the detail and elegance of receptive fields will equip scientists to better evaluate the models against their biological counterparts. Computational models, with the use of receptive fields, can be extended to better purposese. Neural networks of the brain will help scientists understand ourselves and our capabilities and aid doctors in complex medical pathologies.

2 Background

2.1 Layout of the ear

The ear is the earliest stage of auditory processing in the brain. 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 movements of the hair cells allow electrical potentials to travel to the auditory nerve and become processed by the brain.

Hair cells are theorized to be frequency-selective. Not all hair cells respond to all aspects of a sound stimulus. Specific frequencies activate specific areas of the basilar membrane, information which is relayed to higher levels of auditory processing. Many neural networks of the auditory cortex take advantage of this phenomenon by representing sound stimuli as spectrograms, distributions of frequency v. time. Computational neurons in models are connected to a particular frequency range in order to most realistically depict frequency-selectivity.

2.2 Oja's rule

Unsupervised training methods allow neural network models to dynamically modify their own weighted connections between the units, analogous to the changes in synaptic plasticity between real neurons. The simplest form of unsupervised training is based on the Hebbian learning rule. Dr. Hebb hypothesized that if two neurons are simultaneously active, the connection between them should 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 approach infinity after repeated training sets. A modified version of Hebb's rule, Oja's rule, fixes this problem by normalizing the weights during each update 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 paper, the learning rate was arbitrarily chosen in order to avoid these two issues.

Oja's learning rule performs principal components analysis (PCA). PCA determines where the greatest variability of data lies and Oja's rule extracts the first principal component of the data. PCA allows neural networks to meaningfully represent and interpret the input data.

2.3 Spectro-temporal receptive fields (STRFs)

STRFs represent the linear properties of primary auditory processing neurons in many types of animals. 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.

The STRF depicts a unit's impulse response characterizations at frequency-time points. The STRF plots describe the neuronal response in both spectral and temporal terms, and so are more useful than traditional methods of describing neurons such as tuning curves. Tuning curves only depict a neuron's spectral properties. Although STRFs cannot fully capture the properties of an auditory neuron, they are useful descriptions of the cells. Since STRFs can approximate the linear properties of a neuron in both frequency and time, STRFs have been used to predict the outputs of neurons, further validating the utility of receptive fields in the auditory world.

3 Methods

3.1 Neural network

Neurons in the primary auditory cortex have been generally characterized as linear. The neural network employed in this paper is a two-layer, linear transform model. The first layer of the network is the input layer, composed of 129 units. The input to the model are represented as spectrograms in order to account for frequency-selectivity in the basilar membrane. This method of input is the most realistic way of depicting auditory stimuli in the lower levels of processing. The input vector is a single timestep of the spectrogram of the input sound stimulus. Each vector represented a temporal window timestep of approximately 12 ms. The frequencies were evenly spaced by 43 Hz and ranged from 0 to 5512 Hz.

Each neuron in the second layer of the model is linked to 30 input units through the weighted connections. The neurons in the second layer have a 11 input unit window overlap to the two contiguous neurons. In total, the second layer consisted of 6 neurons. The second layer is simply the matrix multiplication of the first layer and the second, or:

$$y_j = \sum_{i=1} w_{ij} x_i \tag{3}$$

where y_j is a second layer neuron, w_{ij} is the weights between the inputs and second layer neuron, and x_i is the first layer neuron. 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 actual connections between the computational units will not change.Through 6 neurons in the second layer may seem small, this allows each neuron to observe a larger frequency range, making the outputs of each individual neuron more meaningful.Each neuron is in the second layer is permanently connected to a set of input neurons.

Animals do not have an unlimited ability to hear all frequencies. In this paper, the neural network has a range of hearing from 214 to 5512 Hz. Although animals may not have the same range of hearing as the neural network has, the neural network artificially simulates this limitation to more accurately match the real world. Therefore, the neural network has no weighted connections between the second layer and the first 5 input units. The upper range of hearing was determined by the sampling rate of the sound stimuli, as the Nyquist Theorem shows that the highest frequency for a sound waveform can accurately capture is half the sampling frequency.

3.2 Temporal coding

The previous neural network does not account for many other properties of auditory neurons. Sound is a function of both frequency and time, and the original neural network only responds to the spectral component of sound. In order to make the computational model more accurate, a simplistic version of memory was also coded into the neural network. Neurons respond to stimuli over time because of neurotransmitter repuptake and degradation; the effect of one neuron on another does not immediately terminate, even after the pre-synaptic neuron is repolarized.The temporal-coded model is able to respond to time-delayed inputs. The previous timesteps are scaled down by factors according to an exponential decay curve, and the sum of all the individual outputs to the time-delayed inputs becomes the response of the neuron. Mathematically, this can be represented as:

$$y_j = \sum_{i=1} \sum_{t=0} \lambda_t w_{ij} x_{i-t} \tag{4}$$

where y_j is the final response, t is the memory span of the model, λ_t is the scaling factors from the exponential decay, x_{i-t} is the previous input, and w_{ij} are the weights.

The different outputs of both networks to a basic, pure-tone stimulus can be visualized in figure *. The model without memory immediately spikes at the onset of the stimulus, remains constant during the stimulus, and immediately returns to zero after the stimulus is over. On the other hand, the model with memory does not immediately increase to its maximum intensity at the onset of the stimulus and returns to zero through an exponential decay.

3.3 Unsupervised training

This neural network model necessitates the use of real-world, natural stimuli. These stimuli were found in CDs, DVDs, and the Internet. The stimuli were batch processed in Adobe Audition 2.0 and altered to a mono, 11025 Hz sampling rate, and 16 bit resolution setting. Although many stimuli were sampled above 11025 Hz, the vast majority of these stimuli had no discernable frequencies above 5000 Hz. As a result, most of those stimuli were composed of just background noise. The sampling rate was reduced the maximum possible frequency at 5512.5 Hz, according to the Nyquist Theorem.

Afterwards, these stimuli were converted into spectrogram matrices (time v. frequency) and scaled between 0 and 1 in Matlab. These matrices served as the input training data to the neural network. After each presentation of a single timestep, the model modified its own weights according to Oja's rule. After each 75 iterations through the training set, the learning rate was annealed by a factor of 10 so that the weights can successfully approach the principal component of the training set. The learning rate was annealed three times, so the model iterated through the training set of 1141 spectrograms a total of 225 times. The weights were saved for further use.

Although Oja's rule changes the weights to represent major features of the natural stimuli, the weights of some neurons were negatively correlated with the regularities of the natural stimuli. These neurons peaked in the negative plane of the ordinate in response to the various environmental sounds. Therefore, the additive inverse of the weights of those specific neurons were used in order for the neurons to best correspond to the natural auditory scenes.

3.4 Constructing stimuli

3.4.1 Moving ripples

The moving ripple stimuli are complex, broadband noises that are used to determine the spectrotemporal receptive fields (STRFs) of neuronal populations. These stimuli analogous to the Gabor patches in the visual domain, both of which have been tested in many previous studies. The moving ripples were made within the same range of frequencies as the natural stimuli used in the training in order for the units in the neural network to respond to their best frequency (BF). The spectral envelope of the noise was then modulated linearly. The ripple equation, intensity at specific time-frequency points, is given as:

$$S(t,x) = 1 + \Delta A \times sin[2\pi(\omega t + \Omega x) + \Phi] \tag{5}$$

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 in Matlab using a program designed by Powen Ru*. 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 40 Hz in steps of 4 Hz and the ripple frequency was varied from -4.6 to 3.4 in steps of 0.4 cycles/octave. The raster outputs of the units to the different moving ripples were computed. The transfer function (TF) is a broad characterization of a unit's responses to the ripple stimuli and is defined by:

$$TF(\omega,\Omega) = M(\omega,\Omega) \times exp[i \times \Phi(\omega,\Omega)] \tag{6}$$

where $M(\omega,\Omega)$ is the response magnitude, $\Phi(\omega,\Omega)$ is the response phase, and $i = \sqrt{-1}$. In order to construct the TF, the magnitude and phase of the raster responses were calculated by performing a Fourier transform. The second half of the Fourier transform was discarded because it provides redundant information. The magnitude was the maximum value of the transform, and the phase was extracted from the unwrapped angle at that point.

The transfer function is assumed to have conjugate symmetry. A two-dimensional inverse Fourier transform function was performed on the transfer function in order to generate the desired STRF.

3.4.2 Tuning curve tones

Tuning curves have been used extensively in both biological and computational applications. Tuning curves allow scientists to quantatatively analyze the frequencies at which a specific auditory neuron responds best to. 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 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 paper, 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 environmental stimuli, and were subsequently converted to spectrograms to become the input of the neural network. The frequency of the tones were varied from 10 to 5490 Hz in steps of 40 Hz. The responses of the computational neurons to these tones were collected. These responses were plotted in a intensity v. time plot, and the peak of the plotted curve denotes the BF of the neuron.

4 Results and Discussion

4.1 Receptive fields

The receptive fields were constructed after probing the neural network with the moving ripple stimuli. The moving ripple stimuli were varied across the rippled velocity (Hz) and ripple frequency (cycles/octave) parameters. The ripple velocity was varied from -40 to 40 Hz in increments of 4 Hz and the ripple frequency was varied from -4.6 to 3.4 cycles/octaves in increments of 0.4 cycles/octave. In total, * ripple stimuli were used to obtain the receptive fields. The other parameters were held constant for all ripple stimuli, shown on table *.

The receptive fields for the six different computational neurons are plotted in figure *. The abscissa represents time after stimulus onset and the ordinate represents frequency. The bright* area of the graph shows where the neuron responds most intensely to. The dark* area shows where neuron does not respond to, or responds very weakly to. The width of the receptive field shows how long the neuron responds to complex stimuli and the length of the receptive field shows the frequency range at which the neuron responds to.

4.1.1 Original model

The STRFs were generated with the neural network without any memory. In the spectral domain of the receptive fields, the graphs show that the neurons are responding to distinct frequency ranges. As the neurons are connected to higher frequency input units, the receptive fields show that the neuron also responds to higher frequencies. This result agrees with the hypothesized outcome and is graphed in figure *. The temporal component of the neurons is constant for all units. Since the neural network is linear, it does not respond to the stimuli over time and so the computational neurons all have the same temporal properties. The temporal domain of the graph is very narrow because the neurons are only responding to one timestep at a time. Additionally, the bright area of intense response is near zero on the time axis. This means the neurons respond immediately after the onset of the stimulus.

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 a neuron responds best to. The BF for all neurons are graphed in figure *.

4.1.2 Tuning curves

After collecting the responses of the neurons to the pure tones, the tuning curves were obtained. The graphs of the tuning curves of all the computational neurons are shown on figure *, in intensity v. frequency. Although biological neurons can only respond at one fixed strength (all-or-none principle), the intensity of the response can be quantized as the firing rate, or how many times a neuron fires per time unit. The intensity of the response in the neural network is analagous to the firing rate in biological neurons. The maximum of the tuning curve is the best frequency (BF) of the neuron; the neuron responds most strongly to frequencies near the BF. The tuning curves, similiar to the receptive fields, show that the neurons respond to specific frequency ranges. The BFs from the STRFs is closely correlated (r=?*) with the BFs from the tuning curves. The tuning curves, though, do not give any indication of how the computational neurons are responding over time.

4.1.3 Model with memory

The receptive fields were also generated from the neural network with a simplistic version of memory. The spectral component of the STRFs from the model with memory is similar to the model without memory. The neurons are still connected to the same frequency ranges, so these STRFs show that the neurons are still responding to the same frequency ranges. Now, the STRFs show that the model with memory is responding over time, evidenced by the repeating areas of high intensity in the images. The subsequent areas of activation in the STRFs are less intense than the original area because of the exponential decay factors that scale the time-delayed inputs.

4.2 Accuracy of models via STRFs

Receptive fields have been used in this project to establish the linear properties of neural networks. Another goal of the project was to help scientists determine the accuracy of neural networks. To this effect, the neural network was modified with different levels of realism by changing the connective pathways between the input and output layer and the values of those connections. The original neural network (NN1) employed structured connections between the input and output layers with weights trained using Oja's Rule. A more unrealistic neural network (NN2) used the same structure, but with untrained weights. The most unrealistic network (NN3) used random connections between the input and the output layers with untrained weights. The receptive fields from each of these networks is shown in figure *. The STRFs show not only visually depict the unrealistic nature of the networks but can also guide scientists in determining the nature of the problems.

The STRFs, though, show nothing new compared to the tuning curves. The tuning curves from each of the neural networks is shown in figure *. The tuning curves show as well how the varying models are unrealistic. But if the temporal-coded model is used, STRFs have the advantage over the tuning curves. For instance, if the neural network with memory is coded with responding to time-delayed inputs with increasing strength, then the STRFs would be able to show the inaccuracies while the tuning curves show no difference. In figure *, the areas of high intensity occur near the end of the receptive field, showing that the time-delayed inputs are more strongly responded to than the current input. On the other hand, the tuning curves look nearly identical in both cases. The STRFs are able to demonstrate the accuracy of neural networks in both the spectral and temporal domains.

4.3 Limitations

In this paper, environmental stimuli were used to modify the weights between the computational neurons. The environmental stimuli, though, are not perfect. For example, the sounds from the Internet and CDs were recorded using different equipment. The quality of the sound varied between the different sources. In addition, the content of the sound varied. Some sounds included large segments of background noise in between bird calls. Some sounds had more background noise than others had. Some of the differences between the sounds were minimzed through resampling and other modifications in order for the input to the neural model to be fairly consistent.

This neural network was constructed to evaluate STRFs in computational models, and so it is only a basic model. The memory coding for the model adds another dimension to the model, but it is not the most accurate method of simulating time-delayed inputs. The neural network also does not model lateral inhibition, the process through which neurons can suppress the activity of other neurons. Lateral inhibition is derived from neurotransmitters such as GABA that inhibit the post-synaptic neuron.

The use of the STRFs has been realistic in this project, with identical use of rippled stimuli to generate the receptive fields. The major differences are the parameters at which the moving ripples were varied across. But this limitation does not pose much of an importance because the STRFs, even across different ripple parameters, are stable and maintain a similar structure.

5 Conclusion

Receptive fields have been extensively analyzed in both animals and neural networks in the visual domain. In the auditory domain, spectro-temporal receptive fields (STRFs) describe both the spectral and temporal aspects of a neuron. The STRFs have been used in many types of animals, but have not been explored in a computational model. This project describes the construction of a neural network of basic auditory processing and the subsequent testing using STRFs. STRFs can be utilized to analyze the properties of computational models of auditory processing in a visual manner. The receptive fields describe how a neuron would generally respond to both the spectral and temporal aspects of sound stimuli. This characterization of the neural network can then assist scientists in determining the accuracy of their models, without reading lines of code or examining multiple outputs from each individual 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 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 doctors in the medical field. The underlying purpose of neural networks is not just to help scientists understand ourselves better but also put that knowledge to use in helping us. The most prevalent auditory disorder is hearing loss, and neural networks may be able to aid doctors determine the cause and spread of hearing loss. This model can approximate the auditory processing, and so can act as a pseudo-subject in an experimental study for hearing loss. Scientists 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, scientists can tell which parts of the ear are affected by the sounds and how both the spectral and temporal properties changed because of the hearing loss. This neural network would even be able to track the spread of hearing loss, and help determine whether the loss to hear certain frequencies can further contribute to hearing loss. Scientists would be exploring the results to this kind of study with a neural network without actually exposing humans to these potentially dangerous sounds.

Spectro-temporal receptive fields are powerful tools that have now been studied in computational models for the first time. This project demonstrates the flexibility and utility of STRFs in visually describing the properties of neurons in an auditory model and determining the accuracy of the model. This analysis suggests that hearing loss can further be studied without affecting humans in an experimental study.