Computational Neuroscience
Dr. Sharba Bandyopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology Kharagpur
Week – 12
Lecture – 57

Lecture 57 : Optimal Coding in Auditory System

Welcome. So we are into our last week and we have started discussing theoretical approaches to neuronal coding. And in the last lecture we talked about how simple cell receptive fields emerge from theoretical principles by applying those principles on a huge set of database of natural images. And the similar kind of ideas that we talked about was extended by Lewicki, Smith and Lewicki and Lewicki. So by Lewicki et al. in around 2003 where they approached the problem of auditory encoding from these theoretical principles.

So as we have been saying here what we will mean by keeping information intact. What we mean by that is okay if we have a sound waveform, let us say it is something like this and we have spike trends in the neural system that we want to be studying and you have a set of them, a number of spike trends across multiple neurons of the neural system in response to that stimulus. What we are saying is that based on these population of spike trends if we can reconstruct the same signal that is if we are going to let us say the red is the reconstruction one and so basically we will try to minimize the error, the difference between the red and the black by what the transformation that we are using in going from the spike trend to the reconstructed signal. So there is some transformation that is happening that is taking these spike trends over onto and creating the red signal and that is again what we will mean by minimizing the error or keeping the information intact.

The second part again that it is sparse coding but that is implemented in this work in a different way. So it does go forward in terms of what was done by Smita and Liviki and Liviki is that so if our x_t is the original sound signal, here we have a dynamic signal earlier in the visual systems work we had static images. In this case we have a temporal component of the sound and of course the pressure amplitude and what we have are kernels that are representing parallel channels of information and they are these ϕ_m's are representative of each of those channels. And so what we have is this s_{im} is essentially the weightage of the ith presence of the mth feature. So ϕ_m is a particular acoustic feature and we have m of them and these are basically like parallel channels.

Those ϕ_m's are parallel channels. So parallel channels and they are each individual acoustic features and these parallel channels have to be weighted by some value every time they are being activated. So in other words at the position τ_i so if we have a signal that is being represented by these acoustic features the mth feature is occurring at the time point τ_i and it is occurring with the amplitude s_{im} and if I can if I get all the i's for all the m's we sum them over and that is the representation this is the representation of the signal x t. So s_i is simply like our activity of each unit only thing is that this activity is present at multiple time points whenever that kind of acoustic feature is present in the sound that is being analyzed and they are present at those particular time points within that channel and the remainder is the residual or the remaining error which needs to be minimized. So they approach the problem of the learning part of these kernels.

So they started out with this kernel ϕ_m to be basically noise, Gaussian white noise sort of with a certain length. So gradually over the learning process as they went over a huge set of natural sounds or speech sounds these ϕ_m's again like what we saw in our earlier work in the visual system this ϕ_m's emerge to be of certain structures and the stopping criteria is obviously based on some minimum error that you have to reach. So the learning process is described here it is simply a gradient ascent on the log likelihood. So if $p(x|\phi)$ is the data probability then they did a gradient ascent for each of the ϕ_m's by taking the derivative with respect to ϕ_m of the log likelihood and that turns out to be simply following this particular expression for a particular kernel m. So by using this gradient and with appropriate steps of learning they used different types of data sets in order to see what are the optimal ϕ_m's or that efficiently encode the auditory signals what are the optimal ϕ_m's that emerge.

So here is the idea that they found is that in the top figure here what we see is representation of frequency of the of each of the capital M channels. So each of the acoustic features that were obtained which are let us say if we think of this particular feature that has occurred at this particular time is a low frequency feature which is represented by this red kind of filter or the kernel and that kernel has a center frequency that is if you take the Fourier transform of that kernel then it has a peak frequency and that frequency is what against which the s_m's are s_{im}'s are plotted in this diagram on the top. So the original signal is provided in gray here that is the original signal and what you have in the reconstruction is this also this gray signal and below is the residual after the system has learned and so what have they learned so this is what they are showing is that the 500 hertz kernel is present in this manner with a particular amplitude the size of that s_{im} is represented by the size of the this mark representing the presence of that acoustic feature at that particular time point. So a few representative kernels are shown as part of the original signal so obviously it is not the overall signal that is being shown that is being reconstructed in the in the colored things there are only four of these kernels at different time different frequencies and time points that have been shown so here they have both the like 20; 3000 hertz kernel present twice with different amplitudes or almost similar amplitudes similarly the next frequency is present here. So these are represented in this blue kernels this high frequency that the smaller amplitude is present embedded in here and so you can now imagine that by adding each of the presence of each of these acoustic features in different amounts throughout the sound duration that is only 25 milliseconds we can get a reconstruction like the grey line below and the difference between them is the residual shown at the bottom trace.

So they have minimized the residual and obtained what are the sizes of s_{im}'s for each of the kernels and that is represented in this particular plot which is called the spike gram. So what they are showing essentially is like a output of a frequency channel if you will and when that channel is active it is active when that particular feature is present in the sound that is this feature is present in the sound so that particular channel is active with a high amount of energy or activity when this feature is present in the sound then we have a high amount of energy in this particular channel with high activity and so on. So now if we consider this to be like a spike but with an analog size or it can be also thought of as a detection of a particular event in the acoustic signal that is this particular acoustic event occurred around that particular time point around this particular time point this particular time point this acoustic event occurred which corresponds to the kernel at 1000 hertz. So now what is what are we seeing in terms of these kernels. So first of all how do we actually when we have a new signal and we have our basis kernels the ϕ_m s how do we know how can we create this spectrum so what they did was essentially a matching pursuit algorithm that is they take the signal in gradual steps and as you move along the signal you first take the first set of the signal project it onto all the ϕ_m's whichever is the largest we take that out as if that particular event is being is occurring within that period of time and then you subtract that out and carry forward this is what happens in matching pursuit and then you get to the next higher kernel the next higher kernel and so on and proceed over in time.

So what are these kernels is how is it related to our actual auditory system. So in order to study that what they did in as they have found the center frequency they essentially found what these kernels are in terms of their frequency representation. So remember this is like the if you have an if you have an auditory white noise stimulus and an auditory nerve fiber is responding to that white noise stimulus and if you do a spike triggered average as you have learned then these kernels are nothing but like the reverse correlation on the spike triggered average obtained for different frequency channels of our auditory nerve fibers in the cat auditory system. So they compared the similar frequency spike triggered averages or reverse correlation functions reverse correlation functions or the spike triggered average obtained from cat experiments cat auditory nerve fibers which are like gamma tone filters or they are very well represented by gamma tone filters and what they find is that the functions or the kernels that they obtained from theoretical principles as we discussed they matched very well with those reverse correlation functions of auditory nerve fibers. So in other words the filtering properties of the auditory nerve fibers so this is what is shown to the left hand side of variety of frequencies low frequency to high frequency is matching with these theoretical kernels and when you look at the bandwidth versus the center frequency of these filters which is what is represent what we call Q factor of a filter that is how sharp the filters are one way of quantifying the sharpness of filters that high bandwidth or low bandwidth depend and based on center frequency what in the auditory system we have is a gradual increase in sharpness as you go higher in frequency that is represented here.

So in general the bandwidth is actually increasing with increasing center frequency and so the Q values can be represented for each of these kernels and overlaid with the auditory nerve fiber Q values or bandwidth versus center frequency values and those are in blue dots for the cat and the red dots represent that for the functions that were derived from natural sounds and as you can see that the span of the blue dots is very much covered by the span of the red dots which are the theoretically derived filters. The black and the green symbols are something else basically they have to do with tonal sounds that are natural tonal sounds and also musical sounds that are present that have a different kind of property and theoretical filters obtained from that other kind of sounds did not match with the cat or red fiber filters which do not need to process that kind of sound. So this was the way that they showed just like Olshausen and field showed that we have emergence of simple cell receptive fields although that is at a much higher level in the sense that it is in the cortex in the primary visual cortex. Here the same kind of filtering from the same theoretical principles is being achieved at the very periphery in the auditory nerve which is the first stage in auditory processing and it is actually carrying information into the central nervous system. So now what they also looked at is basically how it compares with speech and how it compares when we use speech as the sound set.

So instead of natural sounds when they use speech sounds as the database they found another set of kernels that are again shown in red and those are again matching with what you have in actual human kind of auditory nerve indirect measurements of bandwidth and center of frequency from human ears. This is based on auto acoustic emissions. You can get an estimate of the filtering and that is similar to what is obtained in terms of the kernel's properties using the same theoretical principles that we use for the natural sounds. So this is another way of showing that with different kinds of sounds we get different kinds of properties and that also matches with what is actually out there in terms of the filters that in the real world are being used by animals in terms of processing those kinds of sounds. So with these two ideas I hope we have been able to instill in you some interest in approaching these problems from totally theoretical angles as we have discussed.

So again in this case also there has been some follow up work by Lewicki only about higher sparse representations beyond the auditory nerve but again things have not been followed up too much and have not been too successful because we will be needing more sort of ideas in terms of making this learn more and more complex features as we go higher up in the auditory pathway or even in the visual pathway. So we will stop here for these lectures and in the next few sessions we will discuss further on some further theoretical topics and research topics. Thank you.