ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA
CESENA CAMPUS

DEPARTMENT OF ELECTRICAL, ELECTRONIC, AND INFORMATION ENGINEERING – "Guglielmo Marconi" – DEI

SECOND CYCLE DEGREE IN BIOMEDICAL ENGINEERING

CLASS – LM-21

MODELLING AUDIOVISUAL MULTISENSORY INTEGRATION IN ALPHA-GAMMA BANDS

GRADUATION THESIS IN

NEURAL SYSTEMS

SUPERVISOR

PROF. MAURO URSINO

CANDIDATE

SYED SAROSH ALI SHAH

Academic Year 2022-2023

Table of Contents

Abstract

In this work, we have utilized a neural mass model to simulate the audio-visual multisensory integration. The idea is to have a system, equivalent to some neuronal populations, that can perform some behavioral functions and should work in oscillatory conditions similar to EEG (Electroencephalography). We started from the work of (Ursino, Cona and Zavaglia, 2010) which consists of four interconnected neuronal populations (pyramidal neurons, excitatory interneurons, and fast and slow inhibitory interneurons). These four populations make up a cortical column, equivalent to one ROI (Region of Interest), which can produce oscillations at a certain frequency, similar to an EEG frequency band. We are concerned with conscious neural processing and attention which is why our focus is only on two frequency bands: Alpha and Gamma. The gamma rhythm is typically attributed to conscious neural processing while alpha rhythm is associated with attention mechanism i.e., to inhibit the functioning of a certain neural population.

The interconnection of more than one ROI makes up a certain brain area dedicated to specific function. In our model, we have used four brain areas; Two dedicated to unisensory processing (one for auditory and other visual modality), one for multisensory processing, and one for generation of an alpha rhythm, which is used to implement an attention mechanism, to inhibit some external stimuli (either one modality or a portion of space). The unisensory and multisensory areas comprise of 180 ROIs to account for 180 degrees in the azimuthal space, while for alpha rhythm generation we used just one ROI which sends its output to other areas, as desired. The inputs are fed to unisensory areas while the multisensory area receives downstream connection from both. The area dedicated to alpha rhythm generation is connected to all the other three areas. The inputs to both auditory and visual unisensory areas are spatial impulses, constant in time, filtered by respective receptive fields.

The model can perform different behavioral functions. It can solve causal inference problem in the multisensory area in case of dual modality inputs (i.e., one input for each visual and auditory modality). In particular, the multisensory area discriminates if both the inputs have the same cause or not. Furthermore, the unisensory areas can also provide inference about causes in case of multiple unisensory inputs like, for example, two inputs at different azimuthal positions in a single unisensory area. Apart from that, the model can simulate ventriloquism effect which elaborates the bias or shift in perception of auditory and visual

position when the two stimuli are closer to each other in azimuthal space. Our model can only perform one type of ventriloquism which is constant in time but varying in space. Finally, the model simulates attention modulation i.e., the human ability to focus on certain stimuli whilst inhibiting other ambient information. The model can either inhibit one modality completely while keeping the focus on the other or, it can inhibit a certain portion of azimuthal space in both modalities while maintaining the focus on the remaining portion.

Chapter 1 : Introduction

1.1. Overview of the multisensory integration

The interaction between living organisms and their immediate environment relies majorly on the sensory information, presented by the environment, and processed by living beings. In humans, specifically, the brain is the central region to process the sensory inputs which are fed through different modalities such as vision, audition, olfaction, somatosensation etc. The coexistence of various sensory modalities and our ability to process them separately or in conjunction with each other or even substitute one in the absence of any other allows an organism to have the best chances of survival. And what is more is that each of these modalities, tuned to several types of energy, produces a unique perceptual experience that enables a distinct view of outside world (Barry E. Stein and Meredith, 1993).

The interaction between different sensory modalities is referred to "Multisensory Integration" which, as obvious, refers to the interaction between senses and the synthesis of each of their information (Stein and Stanford, 2008). The most frequent way to evaluate multisensory integration is to assess how well the cross-modal stimulus affects the response of an organism compared to individual stimuli. An example would be to assess the organism's magnitude of response or the ability to respond to an event that provides both auditory and visual stimuli compared to when either of them are presented alone. Thus, the functional definition of multisensory integration is measuring statistical significance between the number of impulses elicited by cross-modal stimuli and the number of impulses evoked by either of stimuli individually (Meredith and Stein, 1983).

Multisensory integration can either improve or worsen the response of a neuron which is usually called "Multisensory Enhancement" or "Multisensory Depression". The occurrence of either of these in turn depends on relative physiological salience of the event (Stanford and Stein, 2007). Usually, multisensory enhancement leads to an increased likelihood of detection or initiation of a response to given stimuli. Conversely, in the multisensory depression the likelihood decreases to given stimuli. The speed at which a response can be generated is a significant parameter that aids in the detection of an event through multisensory integration facilitates. Apart from that, multisensory integration can also differ significantly between different neurons when they are exposed to different cross-modal stimulus combinations,

while the variations in magnitude reflect various underlying computations (Stein and Stanford, 2008).

It is worth noting that that enhancement (or depression) is not just “super-additive” but can be sub-additive or simply additive. Super-additive integration is simply the traditionally understood enhancement that is, multisensory response not only exceeds the most vigorous of individual unisensory responses but also exceeds their sum. Additive integration is when multisensory response is simply the sum of individual unisensory responses, while sub-additive integration refers to weaker multisensory response compared to the sum of the individual unisensory responses. Figure 1.1 depicts an example of such behaviors.

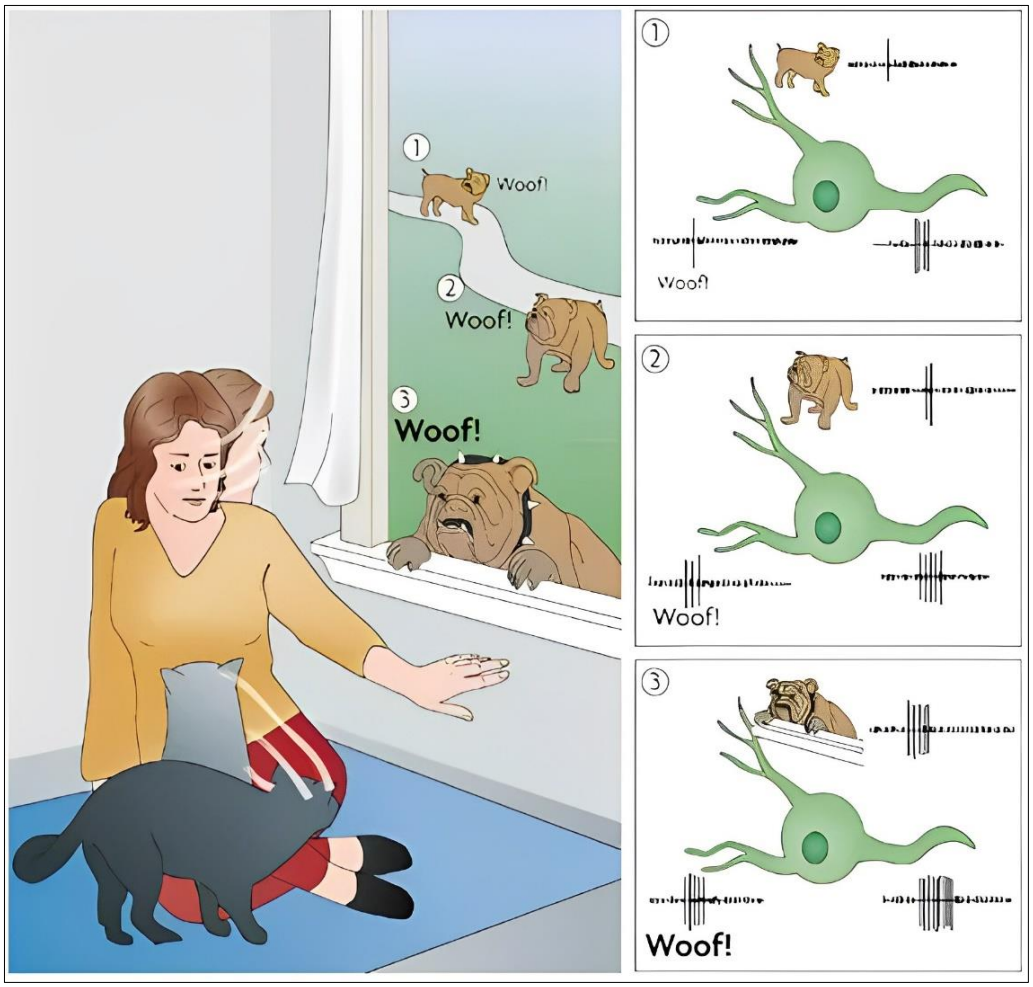

Fig.1.1. Super-additive, Additive, and sub-additive neural responses (Stein and Stanford, 2008).

Figure 1.1 shows how the woman and the cat respond to the incoming dog based on auditory and visual stimuli. As the dog is far away from them, both the stimuli are weak while the integration of senses is super-additive. In the second case the dog is at an optimal position where the induvial stimuli are now stronger, and their integration is simply their sum. The

third case shows subadditivity wherein the multisensory integration is smaller compared to the sum of both individual responses. It is evident that with the increase of individual unisensory stimuli the multisensory integration becomes proportionally smaller.

Thus, concisely, multisensory enhancement is inversely related to individual modality responses/stimuli (Meredith and Stein, 1986). This phenomenon is usually referred to as the principle of "inverse effectiveness". Strong and relevant unisensory stimuli are easily perceived and localized, thus limiting the need for any enhancement whereas weak stimuli tend to evoke neurons at lower rate and therefore the multisensory neuronal responses are subjected to enhancement. This provides a better outcome in terms of detection and localization of any input with faster processing times. Though neurons do seem to follow this principle, nevertheless, recent studies suggest that this is not always the case. Auditory and visual stimuli pertaining to perception of speech are not subjected to inverse effectiveness (Ma *et al.*, 2009).

Another peculiar aspect of multisensory neurons is "causal inference". The brain first needs to understand whether two or more stimuli are generated from a single source or not. This is done based on the sensory similarity between incoming stimuli, for example their spatial or temporal proximity, and thus the brain is capable of producing multisensory enhancement or suppression. Although it is not as simple as it sounds since there are other factors involved like complexity of sensory input, prior knowledge of incoming stimuli and any expectation or likelihood of any stimulus. In case of multi-modal stimulus the causal inference is done explicitly by multisensory neurons while if there are multiple inputs from a unisensory modality then the unisensory neurons can also infer whether the inputs have same cause/source or not.

1.2. Multisensory Illusions

When there are slight spatial or temporal discrepancies in the different unisensory stimuli occurring in close spatial or temporal proximity then a conflict arises in the multisensory integration and thus we often perceive a shifted temporal or spatial event leading to an illusion. There are many illusionary effects known till now, but a few will be discussed here. One common illusion is the "ventriloquism effect" which is commonly attributed to a ventriloquist. The puppeteer is the one talking but since he controls the puppet's lip movements, the observer perceives it as if the puppet were talking. This is due to the spatial

discrepancy in sensory information: Auditory information comes from the puppeteer's mouth while the visual information comes from puppet and since there is only a slight spatial discrepancy between both, the localization based on auditory input is perceived as shifted in space towards the visual one. The reason being stronger visual localization ability compared to its auditory counterpart.

Another effect is the McGurk effect (McGurk and Macdonald, 1976) which is arises during speech perception. Previously, speech perception was regarded as an auditory process, but new evidence suggests visual input also contributes significantly to understanding speech. And it is this influence of visual information that leads to the McGurk effect. For example, when we hear the sound "ba", but the lip movements show "fa" then due to stronger influence of visual processing in the multisensory region, we perceive the word "fa". Thus, the slight sensory conflict between both unisensory stimuli leads to illusionary effect where the stronger/more influential sensory modality takes the lead in our perceptual experience.

Another type of illusion pertains to dominance of auditory dominance on the visual modality in temporal domain. When a visual stimulus is flashed accompanied by an auditory beep the subject perceives both correctly but when the same flash is accompanied by two beeps with less than 100ms temporal gap then the subject perceives two flashes (Shams, L; Kamitani, Y.; Shimojo, 2000). It is worth noting that although the auditory modality dominated in this case, it is misleading to say so. Dominance is neither of the modality nor the stimulus but rather the information that brain can extract from either modality to optimally estimate the spatial or temporal presence of the stimuli (Ernst and Bülthoff, 2004). Finally, it should be noted that only the audio-visual multisensory illusions are detailed here since this thesis is focused on that domain.

1.3.Response time and Evoked potentials

It is generally understood that the processing times for auditory stimuli are faster than the visual stimuli. This difference is usually in the range of 40-60 ms because of faster processing in the inner ear compared to retinal information processing (Barry E. Stein and Meredith, 1993). While this is the case in normal unisensory behavior, nonetheless, in the case of multisensory stimuli the reaction time shortens, becoming comparably lower than individual reaction times to respective stimuli (Rowland *et al.*, 2007). Furthermore, (Diederich and

Colonius, 2004) showed that trimodal (auditory, visual, and tactile) stimuli produced faster reaction compared to bimodal stimuli.

On the neuronal level, the study of evoked potential provides both objective and diagnostic evidence of multisensory processing in the brain. Evoked potentials of multisensory stimuli are summed up to provide a better understanding of inputs while if the unisensory evoked potential is already high in magnitude then the summation in multisensory level is not effective. This phenomenon is consistent with the principle of inverse effectiveness: that is the magnitude of evoked potential increases with the prominence/relevance/importance of stimulus and with increased prominence there is less ambiguity for the humans so their response is quick. This correlation between evoked potential's magnitude and the importance of stimulus provides insights about studying mentally retarded or dyslexic children who have difficulty in processing multisensory stimuli and so their evoked potentials are not properly summed up in multisensory level (Barry E. Stein and Meredith, 1993).

1.4.Neural bases of Multisensory Integration

The thesis focuses only on two sensory modalities: Auditory and Visual, and their interaction and integration in the multisensory region. The visual processing is mostly done by visual cortices that are part of occipital lobe of cerebrum which lies in the posterior region at the very back of brain. Auditory processing on the other hand is done by temporal lobe which has multiple areas dealing with speech synthesis and production, although speech production in general is quite a broad term encompassing multiple regions of brain. Temporal lobe lies on the lateral sides of brain around the region where ears are located.

The traditional view regarding multisensory processing in human brain is that information coming from unisensory areas is fused and processed in another region of the brain. While recent studies show that even unisensory areas are implicated in multisensory processing. The receive inputs from other unisensory areas as well multisensory association areas (Ghazanfar and Schroeder, 2006). It should be noted however that the focus of thesis is on the traditional definition of multisensory processing in the sense that the multisensory neurons, possibly representing a separate multisensory region in brain, are involved in multisensory integration. Therefore, they are placed at a higher hierarchical level in the model receiving inputs from both unisensory areas.

1.5. Electrophysiology of a Neuron

Neuron is a basic building unit of the brain that serves the function of information processing, transmission, and storage. Neurons communicate with each other via synapses that serve as junction between them, and information is transmitted through chemical substances called neurotransmitters. Within a single neuron the information is transmitted by action potentials that are short, stereotyped pulses originating from cell-body of neuron called "Soma" and transmitted via axons that are nerve fibers carrying those electric pulses. A neuron receives multiple connections from other neurons at its dendritic side. The receiving-end neuron is called post-synaptic neuron while the transmitting-end neuron is termed pre-synaptic neuron. Though a single neuron can be both pre or post synaptic depending on whether it is at the receiving or transmitting end.

There exists a potential in the pre-synaptic neuron that is sent towards post-synaptic neuron and when the temporal or spatial summation of all pre-synaptic voltage changes crosses a certain threshold (around -55mv) then the post-synaptic neuron produces an action potential. This potential is generated thanks to the ionic concentration differences between extra-cellular and intra-cellular fluids and the neuronal membrane serves as the boundary between these two. There are ionic channels (or non-gated ionic channels) in the membrane through which ions flow, due to diffusion and electrical gradients, so it is a semi permeable membrane allowing passage of some ions and molecules while blocking others. In essence, the sodium (Na+), potassium (K-), and chloride (Cl-) ions contribute significantly towards the genesis of action potential while many other ions have rather an insignificant effect on overall dynamics of system.

In addition to these channels, there is another way of flow of ions which is Ionic pumps. These pumps, also located in the membrane, serve as active transport mechanism thus allowing passage of ions against the concentration and electric field gradients and so consume energy in form of ATP (Adenosine Triphosphate). The normal concentration gradient is a higher concentration of sodium ions outside the cell compared to inside and similar gradient for chloride ions while the opposite case with potassium ions. Additionally, there are negative proteins inside the cell and thus, at-rest, the potential difference between outside of the cell

with respect to the inside is -65mv (also called resting membrane potential) i.e., the inside of neuron is more negative compared to its outside.

Finally, there is a third type of gate, that allows the passage of ions, called voltage-gated channels. These channels are voltage dependent and open or close depending on the membrane potential. Figure 1.3 represents a neuron membrane showing its selective permeability and membrane potential due to difference in ionic concentrations. Both the ionic channels and ionic pumps can be seen which work based on diffusion + electrical gradients, and active transportation, respectively. For ionic pumps, for every three sodium ions pushed outside the cell two potassium ions are pulled inside.

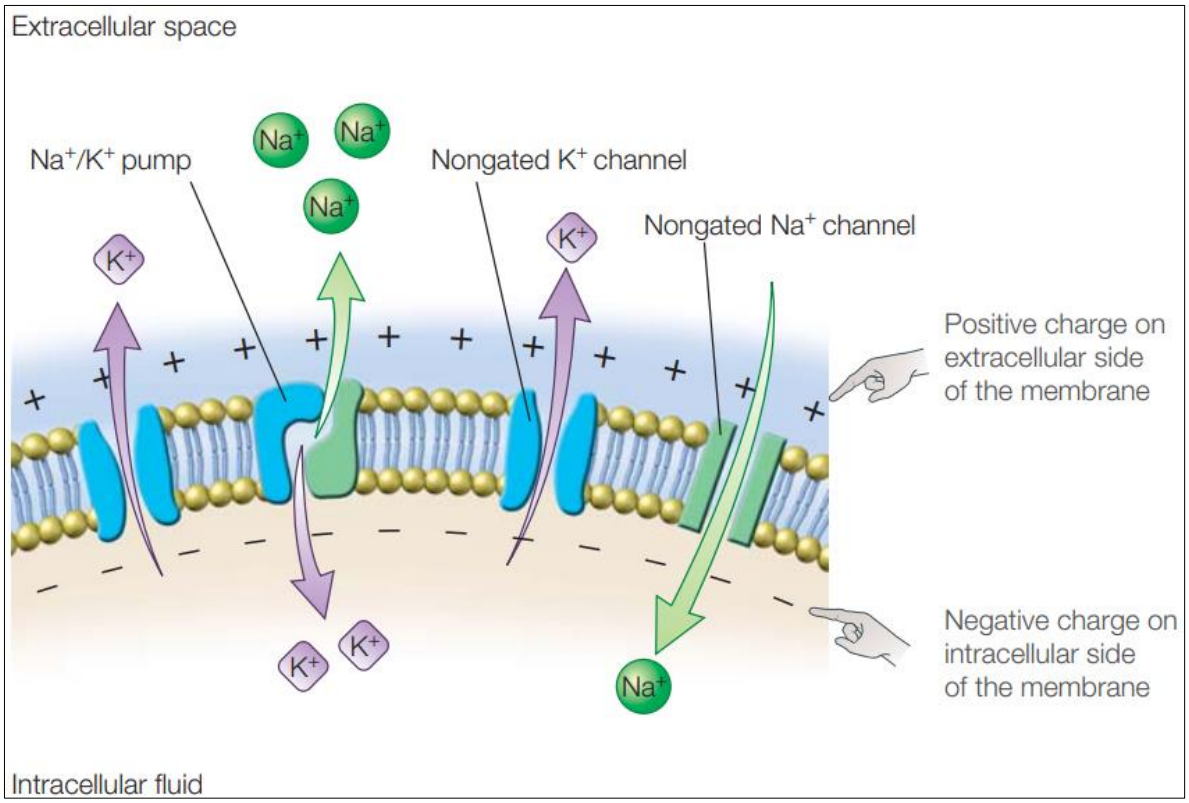

Fig.1.3. Membrane potential of a neuron due to concentration gradient (Gazzaniga, Ivry and Mangun, 2013).

When the post-synaptic potentials, reaching a neuron, sum up and cross the threshold (typically -55mv) then the action potential is generated. Firstly, the sodium gates open allowing it to flow quickly into the cell. The neuron gets further depolarized because of this inflow of positive ions, which further depolarize it and cause opening of more channels. This cycle continues till the maximum number of sodium channels are opened and maximum depolarization is achieved with potential difference reaching around +35mv. It is done in a span of 1ms followed by opening of potassium channels allowing potassium ions to flow outside leading to repolarization or decrease in membrane potential. At this point the sodium

channels have started to close but since the potassium channels outlast the closing of sodium channels the decrement in membrane potential continues beyond the resting potential reaching values of around -90mv. This is called hyperpolarization (Gazzaniga, Ivry and Mangun, 2013).

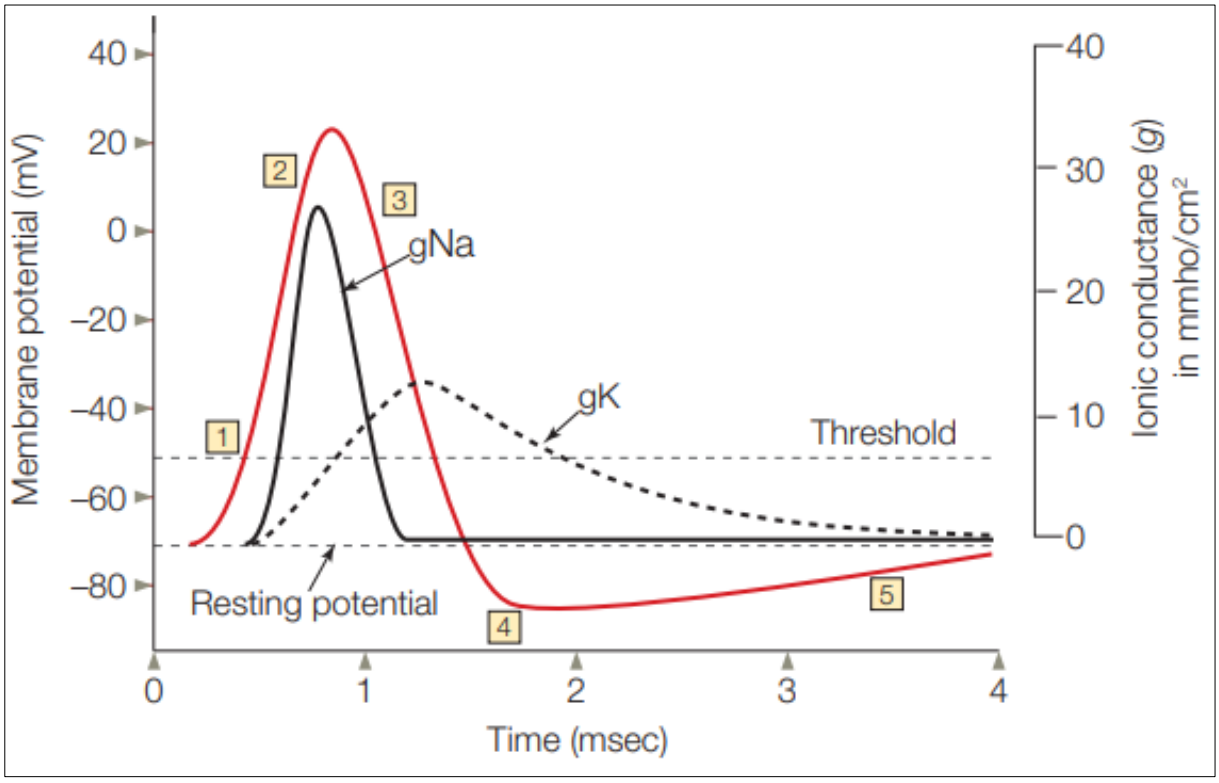

Fig.1.4. Graph of action potential and sodium and potassium conductance (Gazzaniga, Ivry and Mangun, 2013).

Membrane potential during action potential is shown in figure 1.4. which is divided into 5 phases. The first phase shows threshold crossing followed by opening of sodium channels which is also evident in the sodium conductance since it means more sodium flows inside the neuron. Phase 3 shows repolarization followed by hyperpolarization, showing continued outflow of potassium ions. Finally, phase 5 shows membrane potential coming to rest at around -65mv or -70mv showing the closure of potassium channels.

1.6. Computational Models

One of the earliest approaches to model a single neuron based on the action potential was by Hodgkin and Huxley where they used a voltage clamp experiment on squid giant axon. Their analysis led to the famous model of a neuron which can simulate the stereotypical

action potential. What has been stated in point 1.5 and figure 1.4 is represented in figure 1.5 in form of an equivalent electrical circuit. Sodium and potassium conductance are shown as g_{Na} and g_K respectively, while all other ions are merged into one variable called "Eq" which stands for equivalent value. Membrane capacitance is represented as C. Variable conductance are shown that represent ionic pumps as functions of membrane voltage.

The circuit can be modelled as in equation (1). Note that for conciseness conductance are represented as simple variables although they are functions of time and membrane voltage.

$$C\frac{dV}{dt} + g_{Na}(V - V_{Na}) + g_K(V - V_K) + g_{Eq}\left(V - V_{Eq}\right) = i \qquad \textbf{(1.1)}$$

Here the terms g_{Na} and g_K represent sodium and potassium conductance while g_{Eq} represents the equivalent conductance of all other ions. V_{Na} and V_K respectively denote Nernst potentials of sodium and potassium ions, and V_{Eq} is the equivalent Nernst potential of all other ions.

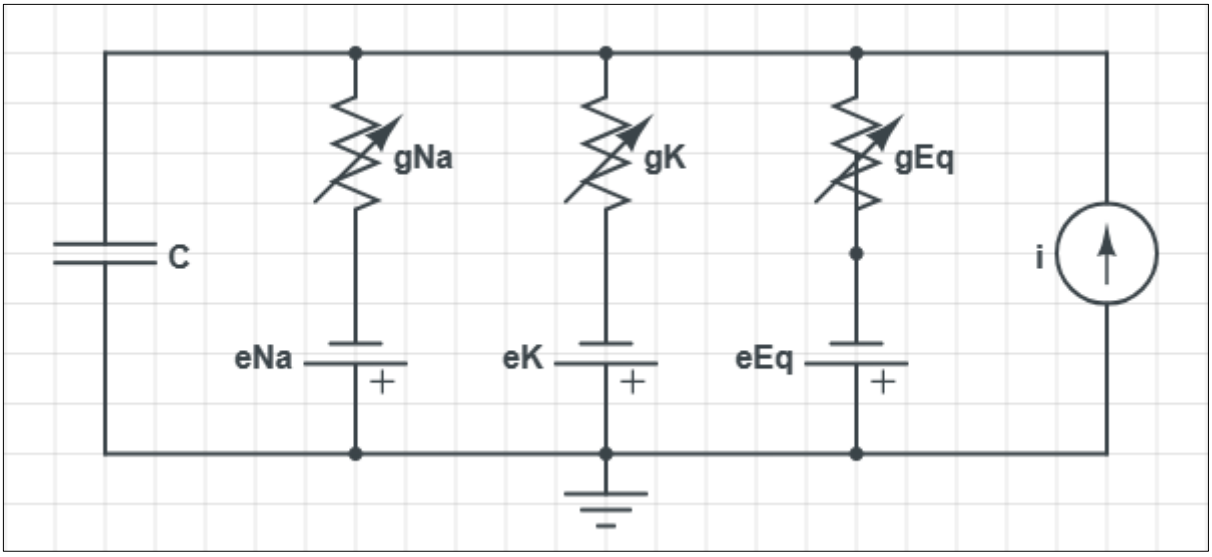

Fig.1.5. Electrical equivalent of Hodgkin and Huxley model.

The current balance equation (1.1) is a concise representation of Hodgkin-Huxley model (detailed explanation can be found in literature). This model is quite robust as it can simulate the action potential in great detail but when it comes to modelling neuronal populations that may contain hundreds or thousands of neurons then the model becomes computationally expensive and therefore a simpler approach is needed. One such model is called Spiking rate model which tries to coarsely capture the stereotypical action potential in form of spikes or impulses. This approach considers the firing frequency as the output of post-synaptic neuron. A simple spiking rate model is shown in figure 1.6.

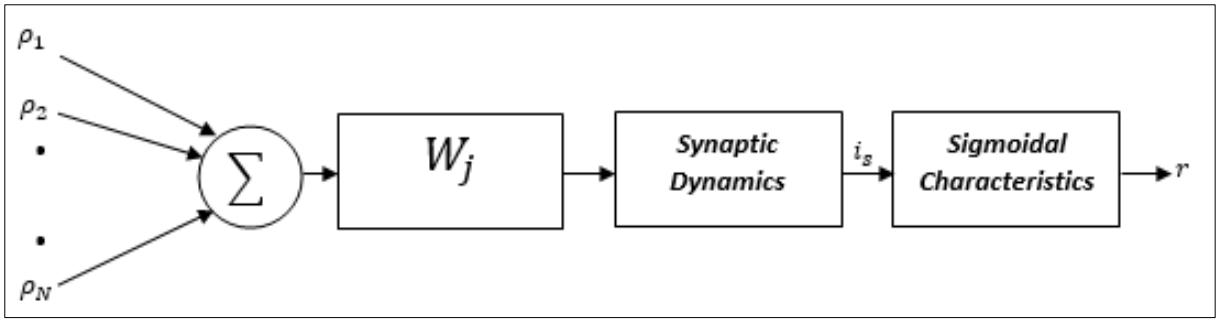

Fig.1.6. Block diagram of a spiking rate neuron model.

N pre-synaptic impulse trains, represented as ρ_j (subscript identifies pre-synaptic neuron's number), sum up and weighted by some factor W_j, and multiplied by the synaptic dynamics to produce the synaptic current i_s which is then passed through a sigmoidal process to provide the rate of firing of post-synaptic neuron. Note that the exact time of impulses is not known a priori thus for simplification, we replace the impulse trains ρ_j with average spike rates r_j.

Regarding this model some important assumptions are made which are as follows,

1. Pre-synaptic neurons have uncorrelated activity.
2. Synaptic dynamics are modelled as linear filter.

Taking these assumptions into account, and assuming a first-order filter for simplicity, the output of post-synaptic neuron can be written as,

$$\tau_s \frac{di_s(t)}{dt} = -i_s(t) + \sum_{j=1}^{N} W_j r_j(t) \qquad \textbf{(1.13)}$$

Here the subscript "s" means post-synaptic neuron while the last term in equation 1.13 denotes the average weighted firing rate which is due to the first assumption. The above equation models the synaptic current which is then fed to a sigmoidal process.

$$S(t) = \frac{S_{max}}{1 + e^{-k(i_s(t) - i_{s0})}} \qquad \textbf{(1.14)}$$

S_{max} is the maximum saturation value when $i_s(t)$ tends to positive infinity while *S(t)* becomes 0 when $i_s(t)$ tends to negative infinity, and the central point is at $S_{max}/2$ when $i_s(t)$ tends to 0.

Again, a concise representation of spiking rate model of neuron is given here. For further reading it is suggested to go through (Gerstner and Kistler, 2002). There are some other approaches to model a single neuron (like integrate and fire model or Fitzhugh-Nagumo model) but are not detailed here since it is evident that using single neuron models to simulate

a population of neurons would require immense computational power. Therefore, an alternative approach is desired that can simulate the macro dynamics of a neural population without providing any detail about their microscopic behavior. One such approach is to use neural mass model which is detailed in chapter 3.

Chapter 2 : Brain Rhythms

2.1. Introduction to Major Frequency Bands

On neuronal level information is encoded as rate of firing of action potentials but it is not enough to tell the complete story about functionality of brain. Typically, a single neuron receives inputs from many other neurons ranging from thousands to hundreds of thousands. While this is the sole means of communication between neurons, on the macro scale it is the post-synaptic potentials that sum up leading to the oscillatory activity of brain. These oscillations are then recorded by measurement systems (Electroencephalography or Magnetoencephalography) to study the brain.

These oscillations are usually divided into five frequency bands.

a. Delta (0.5 - 4 Hz)
b. Theta (4 - 7 Hz)
c. Alpha (8 – 13 Hz)
d. Beta (13 – 30 Hz)
e. Gamma (>30 Hz)

Note that there is slight disagreement between the extreme values of these bands. For example, in some books you might find delta to be in 0.5-3.5 Hz range or Gamma to be in 30-70 Hz range, nevertheless, the differences are quite small. Delta band is usually attributed to brain's activity during deep sleep while theta band is considered to be related to memory encoding and retrieval (Ward, 2003). Alpha rhythm was traditionally considered to indicate cortical inhibition, but recent research shows that it is used as an attention modulation mechanism to inhibit distracting factors while the brain is focused on something. Thus, alpha plays a role in suppressing these unwanted inputs. For example, in memory scanning task the alpha power increases since the brain needs to inhibit any distractions. During the motor imagery task the alpha power is greater on those cortical sites coding for information that needs to be suppressed in order to focus on motor imagery, hence implementing an attention mechanism (Jensen *et al.*, 2002; Cooper *et al.*, 2003).

Beta band is related with an increased conscious activity while gamma band is related to peak conscious performance by the brain. Gamma oscillations have a role in memory processing especially in memory recollection instead of simply experiencing something familiar.

(Burgess and Ali, 2002) observed the presence of higher gamma power in frontal and parietal lobes during recollection task if memory is recalled successfully. They also found that in such tasks, gamma was modulated by hippocampal theta which reflects the coordination between hippocampus and other cortical areas during memory recollection tasks. Gamma band also corresponds to attention mechanism. It has been shown that gamma power is at peak around 300ms after the presentation of stimuli in different sensory modalities, but especially in visual processing (Keil, Gruber and Müller, 2001).

There is also some correlation between gamma band and consciousness. Being visually aware leads to a rise in synchronized firing of neurons in gamma-band. In a study by (Rodriguez *et al.*, 1999), the authors analyzed the EEG signals during shape recognition task where the subjects were shown different images, either containing a face or meaningless shadows. Those who perceived the faces had higher synchronized gamma activity compared to those who did not. Furthermore, (Tononi and Edelman, 1998) suggested that synchronized activity of neuronal firing occurring globally in the brain is the basis for consciousness while localized gamma synchrony is simply unconscious process until it interacts with the global process.

Considering these factors, this thesis is designed to model auditory and visual neuronal populations and their multisensory integration in alpha-gamma bands. In unisensory or multisensory attentive process both populations show gamma oscillations. Alpha band modulation of gamma activity is used either to inhibit one modality, or to inhibit a portion of space, thus realizing an attentive suppression mechanism. Gamma modulated by alpha signifies that the stimulus is still processed, but is inhibited in higher brain areas, since the brain is focusing on the other stimulus.

2.2. Alpha Rhythm Family

Since Hans Berger's research, it has been observed that large, rhythmic alpha waves occur over the visual cortex, particularly when the eyes are closed and there are no eye movements. The appearance of these alpha waves is not triggered by a single cause but is rather a result of certain conditions. The alpha rhythm found in the occipital region is considered a fundamental brain wave. Alpha oscillations are present across the brain and can be weakened by various specific and non-specific stimuli and behaviors. Manipulations like opening the eyes, moving the eyes, engaging in visual imagery, and even mental activities like doing arithmetic

calculations can quickly and consistently block the ongoing alpha oscillations in the occipital region.

Alpha waves are most prominent above the visual areas, but they can also be observed across a large part of the cortex, including the frontal eye fields responsible for eye movement control. The distinct frequencies of alpha waves in these areas demonstrate their relative independence so commonly alpha waves are faster in the occipital region and slower in more anterior regions. The alpha band is defined to have a frequency range between 8 to 13 hertz, although individual variations are substantial. Additionally, the average frequency of alpha waves varies based on factors such as age, gender, and intelligence. Alpha waves in infants are usually found below 7 Hz while peak activity is witnessed in earlier adulthood, and then gradual decrement with age (Buzsáki, 2006).

Usually, these two attributes of alpha rhythm are considered in studies: Alpha power and Alpha frequency. Alpha power, typically pre-stimulus alpha power, is attributed towards excitability level of cortex and accounts for confidence levels too. The excitability level is somewhat a measure of excitability threshold i.e., higher alpha power means higher threshold to excite the cortex on a given stimulus and thus the probability of perceiving a stimulus becomes lower (Ergenoglu *et al.*, 2004). While the alpha frequency accounts for sampling mechanism in the brain: Higher frequency alpha leads to greater accumulation of information is short span of time and thus providing quite accurate perception (Di Gregorio *et al.*, 2022). These mechanisms are studied quite in depth, but the model used in this thesis relies only on alpha amplitude and therefore is restricted to amplitude related functionality.

There is another rhythm that shares a similarity with the occipital alpha in terms of its frequency and, possibly, its underlying mechanisms, however, its occurrence and characteristics differ significantly. Unlike the alpha present in occipital lobe, this rhythm is not influenced by visual input; instead, it is present or produced due immobility of skeletal muscles. Merely making slight movements with a finger or toe can block it. This rhythm is commonly known “mu” rhythm. This motor-relaxation-associated rhythm has been studied by various researchers, leading to the use of different terms to describe it. There is also another type of rhythm like alpha and mu but found in temporal lobe and thus called “tau” rhythm. The basis of this rhythm is desynchronization in middle temporal lobe when subject receives auditory stimuli (Bastarrika-Iriarte and Caballero-Gaudes, 2019).

We have seen a family of alpha rhythms, sharing similar frequency characteristics, however, in subsequent discussion the term “alpha” would be used to denote this whole group of rhythms. Moreover, further focus will be put only on alpha and gamma rhythms and their interactions or influence in multisensory processing.

2.3. Origin of Alpha Waves

Although the exact mechanism of origin of alpha rhythms is still unknown, two major theories attempt to unfold this mystery. One theory explains this phenomenon through pacemaker model, that is, the endogenous oscillations from cortex and thalamus tend to entrain other brain regions and thus alpha activity is produced and propagated. Unlike the first theory that points towards a single group or a few groups of neurons producing alpha oscillations, the other theory makes assumptions on synaptic interactions between neurons that lead to production of alpha waves (Buzsáki, 2006). The former theory is the basis of this thesis since our model makes use of a single neuronal population or region dedicated only for producing and propagating alpha waves.

Alpha waves are dominantly evidenced in occipital lobe of animals who have binocular vision, saccadic eye movements, and have a large visual cortex. Different studies have shown that concurrent electrical recording from occipital lobe and Lateral Geniculate Nucleus (LGN) in thalamus were in sync and thus it was deduced that there was a thalamic role in production of alpha waves. Thus, it is concluded that these waves emerge because of a complex interaction between specific types of neurons in the thalamus (GABAergic thalamic neurons) and cortical regions (thalamocortical neurons), accompanied by the amplification of signals in the neocortex (Buzsáki, 2006).

The prevalence of alpha oscillations in circuits involving the thalamic nuclei indicates that the extent of alpha oscillations reflects how much the cortex is disconnected from processing inputs coming from the body and the environment. However, the relation between this disconnection from certain aspects of the environment and brain performance is not straightforward and so should not be interpreted as such (Buzsáki, 2006). It is also important to note that alpha oscillations are not exclusive to sleep patterns; rather, they are saliently present in the human scalp during various waking conditions (as evidenced from EEG), and not solely due to relaxed muscles or closed eyes, although these conditions do enhance the alpha power.

2.4. Inhibitory Networks in Brain

Inhibitory networks in the brain help maintain stable rhythms. If a network comprises only of excitatory neurons then it will only converge to an irreversible unstable endpoint regardless of the initial condition, unaffected by complexity of synaptic connections. On the other hand, an inhibitory neuron integrated with a chain of excitatory neurons would ensure an overall stable response. In the brain, the pyramidal neurons are excitatory neurons that have long range connections to different regions, whereas short local connections are provided by interneurons that could be both excitatory and inhibitory. Additionally, minute modifications in the parameters of these networks will result in drastic changes in each neuron's activity.

The combination of excitatory and inhibitory elements in networks allows for self-organization and is equipped with complex properties. However, even in the most straightforward connections between these units, the specific wiring details dictate the firing pattern. As depicted in figure 2.1, in a feedback network, increased firing of the pyramidal cell causes the inhibitory interneuron to discharge more frequently which in turn decreases the pyramidal neuron's output. This is a classic example of negative feedback mechanism. On the other hand, in a feedforward inhibitory setup, when the interneuron is activated, it leads to decreased activity of the pyramidal cell. The excitatory input initiates depolarization in the pyramidal cell, but the subsequent feedforward inhibition rapidly repolarizes it, resulting in a narrower temporal window for discharging and consequently the system achieves an impressively precise spike timing of sub-milliseconds range (Buzsáki, 2006).

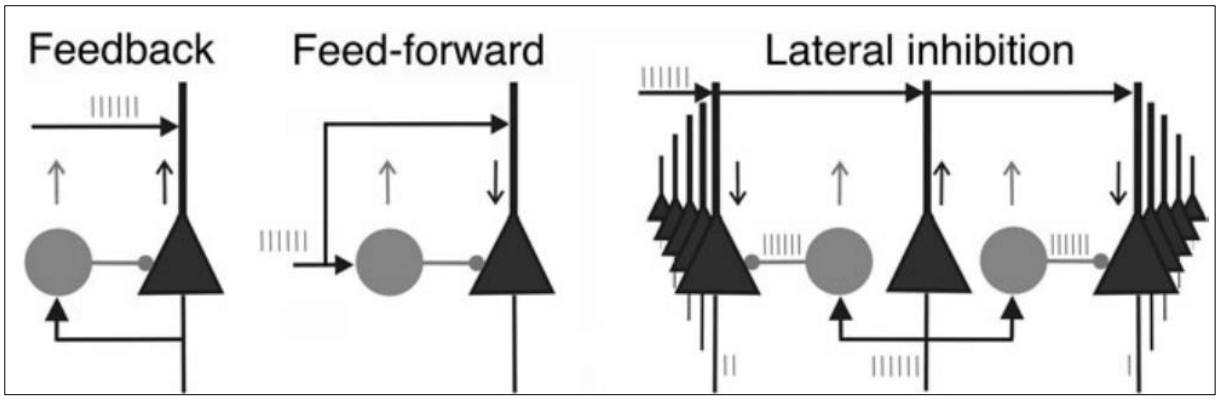

Fig.2.1. Some examples of different topologies of excitatory(black)-inhibitory(grey) networks(Buzsáki, 2006).

When simple feedback or feedforward relationships are altered, the firing patterns of the involved cells become more complex. For instance, when two interneurons are activated

together, their combined impact on the target pyramidal neuron depends on their interaction. In lateral inhibition things become more complicated since it is an extension of a feedback network but with higher complexity. To understand better, consider two pyramidal neurons sharing the same inhibitory interneuron and receiving same inputs. Now in this condition both neurons share common dynamics, but the condition becomes peculiar when one of the pyramidal neurons receives a stronger input. In this case it would tend to suppress the other neuron's activity and the same would happen vice versa. The same situation would occur if the input to any one of the neurons reached slightly earlier in time than the other. Even with such minimal differences between intensities or temporal occurrences of inputs the network's dynamics would differ significantly. It is important to note that the above example can be considered as a "winner takes all" mechanism since any of the neurons receiving stronger or earlier input would define network's behavior.

2.5. Gamma Rhythm

One of the perplexing problems in neuroscience is understanding how the brain puts different pieces of sensory information together to perceive the external world. An apple's shape and color are perceived through vision, its texture is felt through touch so how does the brain integrate this information to perceive an apple? Moreover, the brain is capable of comparing this incoming information to what is already saved inside, and all this is done in the matter of a few hundred milliseconds. This is known as the "binding problem". Another similar problem is superimposition or segmentation problem that pertain to brain's ability to segment different objects that are perceived simultaneously or how it can perceive different objects having superimposed features or patterns such as a salamander and leafy background (Buzsáki, 2006).

One theory that explains this phenomenon is the feed-forward or hierarchical model. This assumes that simple neurons, located lowest in hierarchical level or in-front in feed-forward model, process simple features, but feature processing becomes more and more complex as neurons are placed at a higher hierarchical level. Although this theory got a lot of attention, there are a few problems with it. Firstly, it does not account for physiological feedback connections. Secondly, if such connectionism did exist then for each shape, color, texture, or any attribute or feature, there would be a corresponding neuron encoding that information and a different neuron would be required to code even slight variations in any of the feature.

Ultimately the brain would need almost an infinite number of neurons. Moreover, the brain consists of highly interconnected neurons but if a feed-forward approach is used to model neuronal connections then there would be a dead-end i.e., the last neuron in feed-froward network would have no further connection. Finally, this model is unable to explain how temporal processing is done in the brain.

The alternate hypothesis is of temporal coherence or synchrony in which the temporal synchrony of neuronal firing is crucial. So, when we are perceiving an object, different brain regions, regardless of any "direct" connection and encoding different features, fire synchronously. Such synchronization does not represent causal connectionism as would have been the case for the feedforward model or hierarchical model but provides an acausal description of events temporally bounded together. Such synchrony is typically achieved via feedback connections. It is worth noting that neurons at many times fire randomly, because of continuous changes in post-synaptic potentials, therefore there exists a major problem: To achieve harmonious coexistence of different features, without any error due to random neuronal firing, the brain would need infinite number of connections between different neuronal populations (Buzsáki, 2006).

A solution to this problem is the presence of oscillations which can synchronize in a few cycles leading to effective perception: for example, distinguishing between foreground and background. Such oscillations can function even in sparse connections. The earliest evidence in this regard was provided by (Gray *et al.*, 1989). The authors measured local field potentials in cat's visual cortex and found the presence of gamma band activity due to moving bars. In another study by (Burgess and Ali, 2002), the authors concluded that gamma oscillations arising from neurons in separate cortical columns can synchronize and thus are related to feature binding. For such type of binding it is necessary that the receptive fields of different neurons overlap otherwise the oscillations do not synchronize.

2.5.1. Gamma Rhythm Generation

The interneurons, which usually release the inhibitory neurotransmitter GABA (gamma-aminobutyric acid), are quite less in number in the cortical region and only make up one-fifth of overall cortical neuronal population, or even less (Buzsáki, 2006). Given this scenario, how does this small inhibitory population match up with the large excitatory pyramidal population? There are various approaches through which this small group keeps the

larger group in check. Firstly, the synaptic connectivity between interneurons-pyramidal neurons (around 5-15 synaptic connections between single interneuron and a pyramidal cell) is quite stronger compared to pyramidal-pyramidal connections. Secondly, the activation threshold for an interneuron is low and at times even a single action potential from a pre-synaptic pyramidal neuron can cause the post-synaptic interneuron to fire (Buzsáki, 2006).

Through these mechanisms the interneurons fire more rapidly than pyramidal cells and thus the overall effect of Inhibitory Postsynaptic Potentials (IPSPs), converging on a pyramidal neuron, becomes equal to Excitatory Postsynaptic Potentials (EPSPs). IPSPs are characterized by short rise and decay times and higher amplitudes than EPSPs which have lower amplitude but with longer temporal dynamics. Moreover, the IPSPs normally converge on cell body of post-synaptic neuron as compared to EPSPs which dominantly localize on dendrites of post-synaptic cells. Consequently, the extracellular space, with high density of cell bodies or soma, is filled with high frequency currents in contrast to higher dendritic density areas. Thus, it is this connectivity and interplay of excitatory-inhibitory neurons that leads to cortical harmony and the global firing rates of neurons are kept in check. Furthermore, through such connectionism the excitability threshold of neurons can be lowered for brief periods of time allowing efficient message transmission and network modification (Buzsáki, 2006).

The combined effects of these interneurons then exhibit gamma rhythm. It has been shown that inhibitory interneurons with fast dynamics (denoted as $GABA_{A,fast}$) and slow dynamics (denoted by $GABA_{A,slow}$) are crucial in maintaining different rhythms. Typically, $GABA_{A,fast}$ are needed for gamma generation and $GABA_{A,slow}$ for slower rhythm generation, like theta (White *et al.*, 2000). Note that the subscript *A* here only means that they are *type-A* GABA. Nevertheless, the neuronal dynamics that lead to generation of different rhythms do not simply work standalone but rather there is constant activity in the brain and therefore the rhythms are not generated as separate entities but as a complex time series made up of spectrum of different frequencies contributing to overall brain function. For instance, the alpha-gamma rhythms are attributed to attention: Gamma is associated with conscious signal processing while alpha is related to focus or attention. In particular, a gamma rhythm modulated by alpha is implicated in subconscious stimulus processing in conditions when some stimuli are inhibited by brain to focus attention on something else. Moreover, a gamma rhythm modulated by a slower theta rhythm is implicated in learning and memory (Ward, 2003).

Chapter 3 : Neural Mass Models

3.1. Why Neural Mass Models?

In the first chapter some of the computational models of single neurons were presented which can be used as basic units when modelling population dynamics. These single neuron models are quite detailed in terms of microscopic behavior of brain, that is, they explain the spiking dynamics of neurons well, and allow better understanding of spike timing, propagation of action potential, and synaptic interactions. Through this approach population models can be made more biologically plausible and the contribution of each single neuron towards overall dynamics of the population can be analyzed. Furthermore, very complex connections between different populations of neurons can be modelled starting from the spiking model of single neuron and the population models made from these single neuron models can effectively capture both micro and macroscopic behavior of brain regions.

These models are quite robust in terms of mimicking neurophysiology but are quite computationally expensive, especially when creating very large networks containing hundreds of thousands of neurons, because the parameters in such models could number in the range of millions or even billions and parameter estimation becomes extremely difficult given that they need to be biologically relevant. So, these models tend to be very complex, and their analysis becomes a demanding job. Therefore, a different approach can be utilized. Instead of starting from single neuron model and working all the way up to a population, we can consider the population as a whole and model it directly without considering what happens on microscopic level. This approach may not seem biologically relevant, but on macroscopic scale it can simulate the behavior of a neuronal population.

There have been numerous approaches towards these population models (also called Neural Mass Models or NMMs) since the 1970s. Earlier work by (Wilson and Cowan, 1972) was quite important as it laid the framework for researchers to work on coarse-grained models. Further work by (Lopes da Silva *et al.*, 1974; Rotterdam *et al.*, 1982) provided better insights to simulate alpha band activity. Subsequently, the model by (Jansen and Rit, 1995) gave a better physiological meaning to the computational model. They included three neuronal populations in their model (pyramidal cells that are excitatory and two groups of interneurons, both excitatory and inhibitory) and could simulate the different frequency rhythms but only by changing the synaptic connections and time constants. However, a ROI (region of interest) at

each time could only simulate a single type of rhythm. It was shown by (Grimbert and Faugeras, 2006) that the model was prone to instability so there was a strong limitation on the choice of parameters. Note that we have used the word “population” or “group” to refer to a single type of neuronal network (pyramidal cells or Interneurons) and the word “ROI” to refer to the whole structure (like a cortical column) of these interconnected groups or populations that can jointly generate one or more rhythms.

An extension of Jansen and Rit model was done by (Wendling *et al.*, 2002) wherein they used both fast and slow inhibitory interneurons and simulated epileptic activity. The physiological reality of having multiple rhythms coming from a single ROI was computationally possible thanks to the work by (David and Friston, 2003) where they generalized the Jansen and Rit model and included multiple populations to achieve this milestone. Similar study was done by (Sotero *et al.*, 2007; Ursino *et al.*, 2007) to generate multiple rhythms using multiple populations but with one population producing one rhythm only. Trying to come up with a way of having a population capable of generating more than one rhythms the study by (Zavaglia *et al.*, 2008) made it possible. They used similar model as (Wendling *et al.*, 2002) with 4 interconnected neuronal groups making up one ROI. By connecting three ROIs they were able to simulate two rhythms in a single ROI, with the other rhythms coming from other sources like a different ROI or an external oscillator.

It is now evident that the computational model of different interconnected neuronal populations is able to generate a rhythm and the interconnected populations can work jointly to simulate one or more rhythms. The significant achievement of having one type of population capable of producing high frequency gamma rhythms was by (Jefferys, Traub and Whittington, 1996). In their work they simulated 40 Hz gamma rhythm using only inhibitory interneurons but with self-loop and an excitatory external input. However, the model is quite sensitive to external input’s oscillations and slight changes can cause unstable behavior. Following this work, the authors in (Ursino, Cona and Zavaglia, 2010) modelled a system with four ROIs and each ROI containing four neuronal populations (pyramidal, excitatory interneurons, $GABA_{A,fast}$, and $GABA_{A,slow}$ interneurons). Moreover, the fast inhibitory interneurons within each ROI receive a self-loop which provides better oscillatory behavior in the gamma range.

This latest description is the basis of this thesis, and the model will be presented in detail in the next chapter while the current chapter is dedicated to detail some historically significant

mass models, starting with Wilson-Cowan oscillator followed by Jansen and Rit model, and then the analysis of inhibitory interneuron population with self-loop.

3.2. Wilson-Cowan Model

The Wilson-Cowan model (Wilson and Cowan, 1972) is a mathematical description of interaction between excitatory and inhibitory populations. Its major goal is to simulate the macroscopic behavior of neuronal networks with no attention to detail on singular neuronal contributions. It was a significant milestone in computational neuroscience since researchers were struggling to find ways to model neural dynamics on a global scale. Their model is based on coarse-grained approach meaning that the system is simplified in terms of degrees of freedom so that there is a trade-off between complexity of model and its ability to effectively simulate the system's dynamics. The way forward is to have a simpler model with less degrees of freedom and the information loss is acceptable such that the phenomenon under analysis can be studied optimally. Their coarse-grained approach makes use of mean-field equations that represent the proportion of active neurons, at a given moment in time, in a neuronal population.

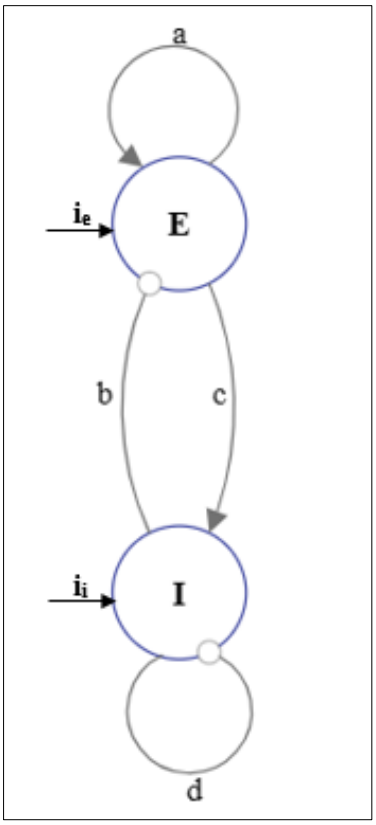

Fig.3.1. Pictorial depiction of Wilson-Cowal model.

Figure 3.1 elaborates the interaction of excitatory-inhibitory populations of neurons based on the Wilson-cowan model. The excitatory population is represented with **E** and inhibitory population with **I**, while *a*, *b*, *c*, *d* are constants showing synaptic strength, and lines connected by arrowhead are excitatory connections whereas circular shape at end represent inhibitory connection. Note that both populations receive self-connections too (*a* and *d*) which does not mean that the neurons excite or inhibit themselves; rather these represent interconnections between neurons in the same population. Finally, the inputs are denoted by i_e and i_i respectively, for excitatory and inhibitory populations, that are considered to be sum of currents incoming from other neuronal populations or external sources.

The model equations can thus be written as (Kilpatrick, 2013),

$$\tau_e \frac{dE(t)}{dt} = -E(t) + [1 - r_e E(t)] f_e(aE(t) - bI(t) + i_e(t)) \quad \textbf{(3.1)}$$

$$\tau_i \frac{dI(t)}{dt} = -I(t) + [1 - r_i I(t)] f_i(cE(t) - dI(t) + i_i(t)) \quad \textbf{(3.2)}$$

These are first order differential equations with time constants τ_e and τ_i, and *E(t)* and *I(t)* show the proportion of active neuronal populations at time instant *t*. Furthermore, the inputs to each population are time varying, which, for reasons of simplifications, can be assumed to be constant in some cases. The synaptic strength factors are crucial in determining the oscillatory activity of model and are carefully selected to mimic physiological values. f_i and f_e are nonlinear activation functions (equations **3.3** and **3.4**) to show the proportion of active of inactive neurons. Typically, a sigmoidal function is used here with 0 meaning complete inhibition and 1 meaning complete excitation.

$$f_j(x) = \frac{1}{1 + e^{-\gamma_j(x - x_j)}} \qquad j \in \{e, i\} \quad \textbf{(3.3)}$$

The factor γ_j is gain and x_j is threshold depending on population type *j* (*e* or *i*). Before the nonlinear function there is a factor *[1-r_jJ(t)] (J∈{E,I}, j∈{e,i})* which defines the refractory period of these populations whereby the neurons are unable to fire, immediately following activation. In some subsequent studies this factor is often neglected but for sake of completeness it is reported here. Analyzing these equations regarding their stability we keep $r_j = 0$, so the equations for isocline are,

$$E(t) = f_e(aE(t) - bI(t) + i_e(t)) \quad \textbf{(3.4)}$$

$$I(t) = f_i(cE(t) - dI(t) + i_i(t)) \quad \textbf{(3.5)}$$

Equation 3.3 is vertical isocline and 3.4 is horizontal. Keeping a constant current input for the excitatory population while 0 input for inhibitory, and setting some constant values for synaptic strengths, the resultant phase plane can be plotted as shown in figure 3.2.

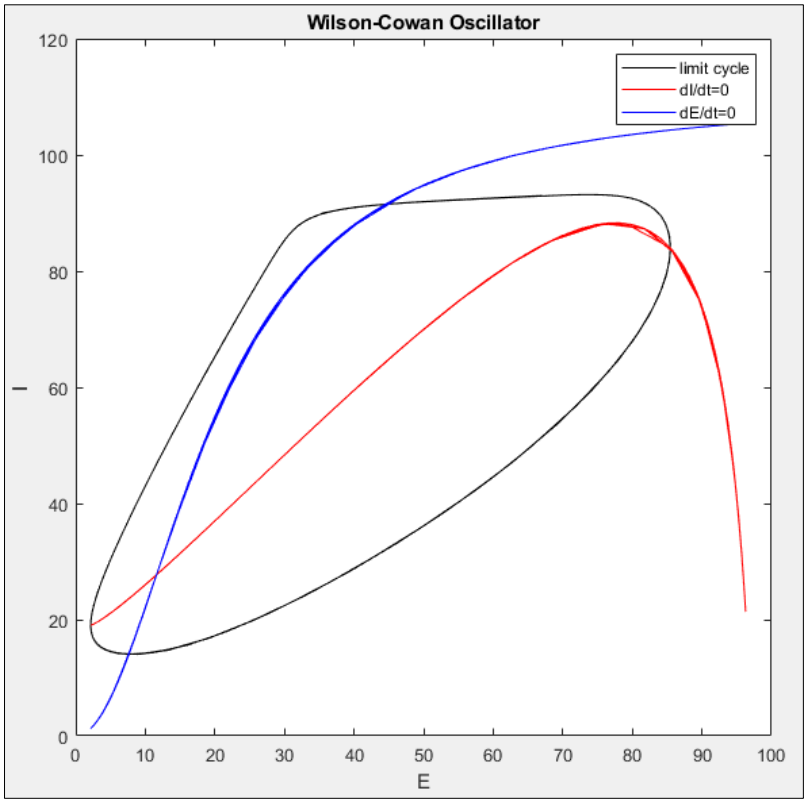

Fig.3.2. Phase plane plot of Wilson-Cowan model..

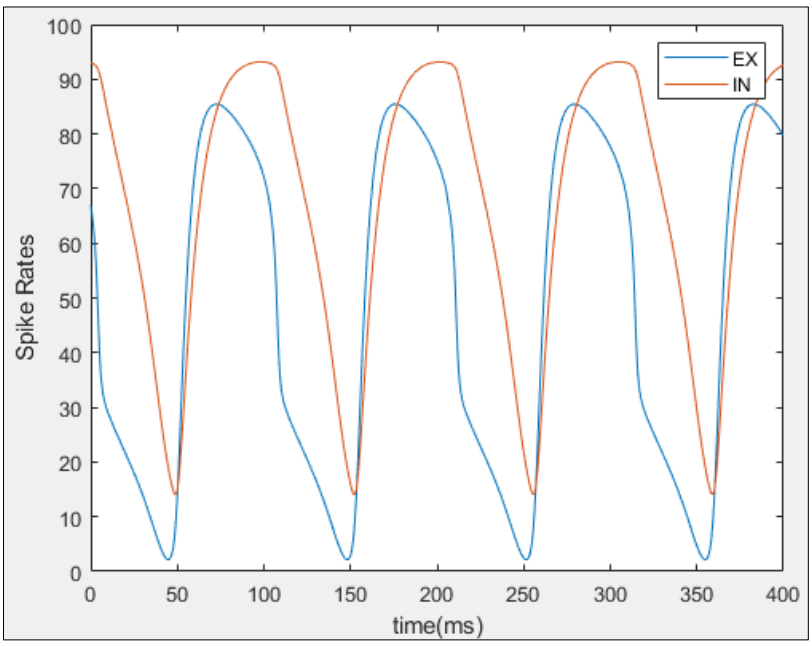

Fig.3.3. Simulation of Wilson-Cowan model.

The values for this plot are taken from (Wilson, 1999) and simulated in MATLAB software. Through further analysis it can be shown that this system provides an asymptotically stable limit cycle describing the oscillatory nature of the model as depicted in figure 3.3. The existence of a limit cycle is restricted to significantly high values of input current and of course carefully selected synaptic strength values. In fact, if the current is too low then a stable equilibrium point would occur where the population activity would converge. Apart from that, to achieve different oscillatory bands the different values of time constants can be used.

The model was initially developed for visual processes but then different studies extended it towards other sensory-motor and cognitive processes (Kilpatrick, 2013). Though this model provides a good analysis on oscillations, it is not quite physiological as its parameters have no relevant physiological meaning. Furthermore, it takes into account only the two neuronal populations (excitatory and inhibitory) in a certain brain region while there are a number of different types of neurons present in a cortical column. Moreover, the model assumes that a certain region of the brain can only have one rhythm at a time while that is not the case. Another simplification is the absence of a clear description of synaptic dynamics. Therefore, a better approach is needed that can account for these limitations.

3.3. Jansen and Rit Model

The Jansen and Rit model is also a coarse-grained or lumped parameter model which simulates a cortical column. The excitatory pyramidal neurons are in feedforward condition whilst receiving feedback connections from both excitatory and inhibitory interneurons. These interneurons can represent local or distant neurons feeding the inputs to pyramidal neurons. The neuronal population in each group is modelled by having two stages of processing: One for transforming the incoming average spike rate (average density of action potentials) into post-synaptic potentials (PSPs) and the other stage for transforming these PSPs back to average density of action potentials (Jansen and Rit, 1995). It is worth noting that the PSPs can be inhibitory (IPSPs) or excitatory (EPSPs), as a consequence of GABAergic or Glutamate based synapses.

Th first stage of processing refers to sigmoidal relationship that transform PSPs (denoted by y_i where i denotes post-synaptic population) to average density of action potentials or simply the spike rate (denoted by z_i) of that population. Then the second stage of processing refers to

synaptic dynamics of a neuron in physiological reality. It is modelled by a second order differential equation. A simple single population model is shown in figure 3.4. The incoming post-synaptic potentials are weighted, summed up (positive or negative depending on excitatory or inhibitory synapses, respectively), and passed through a sigmoid to be transformed into spike rate. Then, the second order synaptic dynamics transform the spike rate to post-synaptic potential.

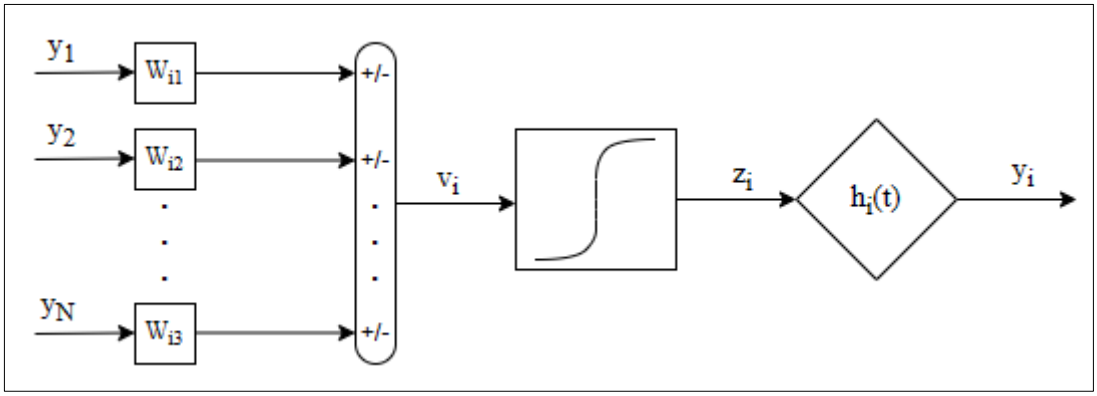

Fig.3.4. Model of single population based on Jansen and Rit.

Following the above representation, the equations for single population can be written as,

$$v_i = \sum_j W_{ij} y_j\,(t) \tag{3.6}$$

$$z_i(t) = S\big(v_i(t)\big) = \frac{2e_0}{1 + e^{-r(v_i(t) - v_0)}} - e_0 \tag{3.7}$$

$$\ddot{y}_i(t) = G_i \omega_i z_i(t) - 2\omega_i \dot{y}_{\,i}(t) - \omega_i^2 y_i(t) \tag{3.8}$$

Here the subscript j represents presynaptic population while i represents post-synaptic population. W_{ij} are the connectivity weights between pre and post synaptic populations and $v_i(t)$ is the weighted sum of PSPs of pre-synaptic populations, while y_i is the PSP of post-synaptic population. The value e_0 is the maximum possible spike rate, r is related with the steepness of sigmoid function and, v_0 is the PSP at which spike rate is 50% of overall value. G_i is the gain and ω_i is the inverse of time constant of the system. In the Laplace domain the synapse equations become,

$$H_i(s) = \frac{G_i \omega_i}{(s + \omega_i)^2} \tag{3.9}$$

$$h_i(t) = G_i \omega_i t e^{-t\omega_i} \tag{3.10}$$

Following this single population model the schema for complete model is shown in figure 3.5 which contains three populations: Pyramidal neurons, Excitatory Interneurons, and Inhibitory

Interneurons. These interconnections follow following typical physiological rules, as defined previously in topic 2.4.

a. Pyramidal neurons do not self-excite but do excite both excitatory and inhibitory interneurons.
b. Excitatory interneurons only excite the pyramidal neurons.
c. Inhibitory interneurons only inhibit the pyramidal neurons.
d. External inputs (from pyramidal neurons in different cortical columns) can excite all three of these populations.

For clarity of notations, the first subscript will be used for post-synaptic population while second subscript will be used for pre-synaptic population, for instance, W_{ep} means the weight for PSP of pyramidal neurons (pre-synaptic) going towards excitatory interneurons (post-synaptic). Similar notations are used for the rest of the variables as in equations 3.6-3.8, except that subscripts *p*, *e*, and *i* represent pyramidal, excitatory interneurons, and inhibitory interneurons, respectively, while n is the input. For sake of simplicity only one input is included in the excitatory interneurons' population.

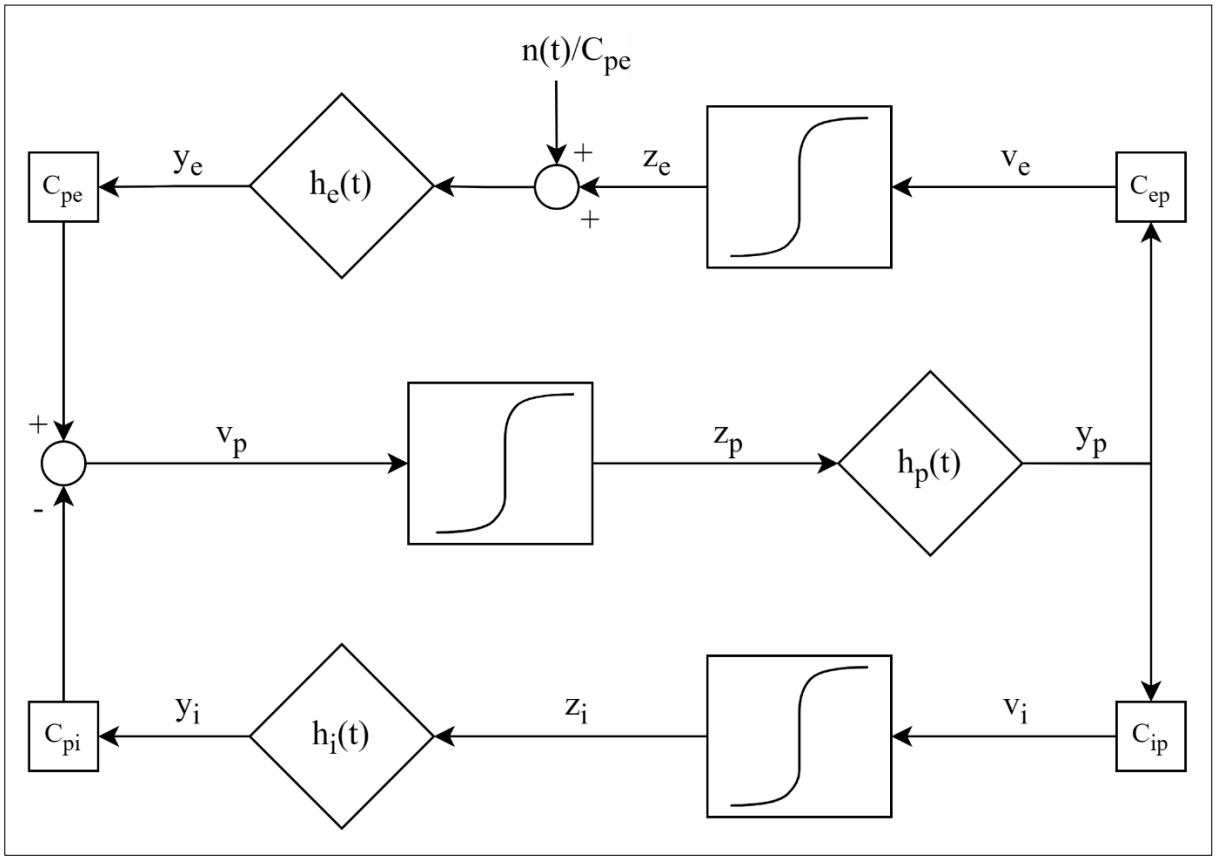

Fig. 3.5. Jansen and Rit model of three interconnected populations.

Note that from now onwards the weights W_{ij} will be replaced by C_{ep} since in the next chapter the notation W would be used for other purposes. The equations for above schema can now be written as,

Pyramidal Neurons

$$v_p(t) = C_{pe} y_{e(t)} - C_{pi} y_i(t) \quad \textbf{(3.11a)}$$

$$z_p(t) = S\left(v_p(t)\right) = \frac{2e_0}{1 + e^{-r(v_p(t) - v_0)}} - e_0 \quad \textbf{(3.11b)}$$

$$\ddot{y_p}(t) = G_p \omega_p z_p(t) - 2\omega_p \dot{y_p}(t) - \omega_p^2 y_p(t) \quad \textbf{(3.11c)}$$

Excitatory Interneurons

$$v_e(t) = C_{ep} y_p(t) \quad \textbf{(3.12a)}$$

$$z_e(t) = S\big(v_e(t)\big) = \frac{2e_0}{1 + e^{-r(v_e(t) - v_0)}} - e_0 \quad \textbf{(3.12b)}$$

$$\ddot{y_e}(t) = G_e \omega_e \left[z_e(t) + \frac{n(t)}{C_{pe}} \right] - 2\omega_e \dot{y_e}(t) - \omega_e^2 y_e(t) \quad \textbf{(3.12c)}$$

Inhibitory Interneurons

$$v_i(t) = C_{ip} y_p(t) \quad \textbf{(3.13a)}$$

$$z_i(t) = S\big(v_i(t)\big) = \frac{2e_0}{1 + e^{-r(v_i(t) - v_0)}} - e_0 \quad \textbf{(3.13b)}$$

$$\ddot{y_i}(t) = G_i \omega_i z_i(t) - 2\omega_i \dot{y_i}(t) - \omega_i^2 y_i(t) \quad \textbf{(3.13c)}$$

To get the oscillatory activity the following parameter values were used (Jansen and Rit, 1995).

Parameter	Detail	Value
C_{ep}	Average synaptic gain from Pyramidal population to Excitatory interneurons' population	135
C_{pe}	Average synaptic gain from Excitatory interneurons' population to Pyramidal population	0.8*135
C_{ip}	Average synaptic gain from Pyramidal population to Inhibitory interneurons' population	0.25*135
C_{pi}	Average synaptic gain from Inhibitory interneurons' population to Pyramidal population	0.25*135
G_e , G_p	Excitatory Synaptic Gain (Same of Pyramidal neurons and Excitatory Interneurons)	3.25 mV

G_i	Inhibitory Synaptic Gain	22 mV
v_0	PSP at which 50% spike rate achieved	6 mV
e_0	Maximum Possible spike rate	2.5 s^{-1}
r	Steepness of sigmoid slope	0.56 mV^{-1}
ω_e , ω_p	Time constants of pyramidal neurons and excitatory interneurons	100 s^{-1}
ω_i	Time constant of inhibitory interneurons	50 s^{-1}

Table 3.1. Parameter values for Jansen and Rit model.

This model has been shown to be quite useful in simulating oscillatory behavior of neuronal populations in different frequency bands, especially lower frequency bands like delta, alpha or beta. It has been employed to simulate EEG/MEG brain activity (David and Friston, 2003), model epileptic conditions (Wendling *et al.*, 2000), or to study brain dynamics using Dynamic Causal Modelling (DCM) (David *et al.*, 2006). The model is quite robust and able to simulate different conditions pertaining to brain rhythms.

3.4. Inhibitory Interneurons with Self-Loop

In the previous mass models, populations in each ROI were either interconnected and thus received synaptic inputs from each other or received long range inputs from pyramidal neurons of other ROIs (Jansen and Rit, 1995; Wendling *et al.*, 2002), and so the oscillatory activity was ensured. Conversely, a population simulated only via a white noise input could not achieve rhythmic behavior (Ursino, Cona and Zavaglia, 2010). But however, different studies evidenced that fast inhibitory interneurons can solely generate gamma rhythm as they are highly interconnected with each other (Bartos, Vida and Jonas, 2007). Computational studies in this regard have shown the same i.e., gamma rhythm can be generated by fast inhibitory interneurons alone without any influence of other neuronal populations (Jefferys, Traub and Whittington, 1996; Sotero *et al.*, 2007).

Following these approaches, a simplistic model of these interneurons is detailed in figure 3.6 (Ursino, Cona and Zavaglia, 2010) that shows a self-loop along with any possible external input (from neural population within same ROI or pyramidal cells from other ROIs). This inhibitory population is the one with fast dynamics ($GABA_{A,fast}$) and the input is simply white noise with zero mean. To analyze this system in an isolated manner the external input is kept zero and the non-linear block is linearized around the critical points of state vector Y_f.

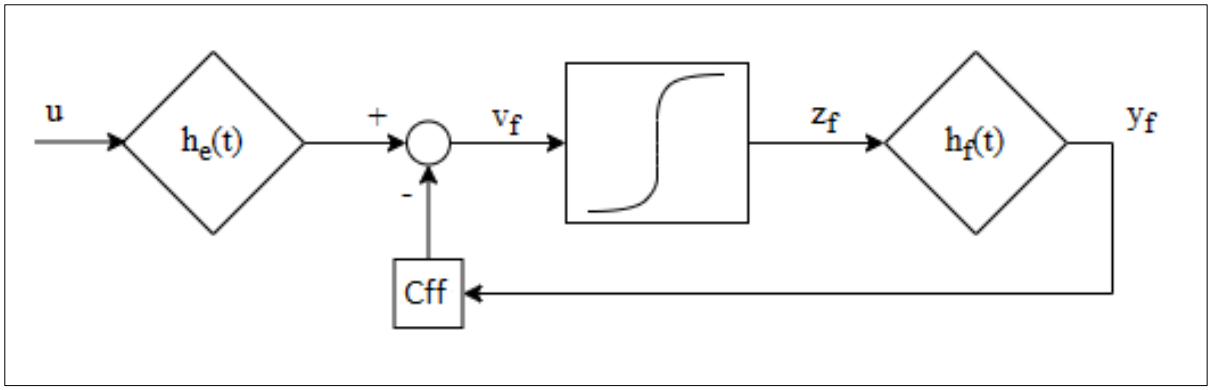

Fig.3.6. Fast inhibitory interneuron population with self-loop.

The equations are thus written as,

$$v_f(t) = C_{ff} y_f(t) \quad \textbf{(3.14a)}$$

$$z_f(t) = S\big(v_f(t)\big) = \frac{2e_0}{1 + e^{-r(v_f(t) - v_0)}} - e_0 \quad \textbf{(3.14b)}$$

$$\dot{y_f} = 0 \quad \textbf{(3.14c)}$$

$$\ddot{y_f}(t) = G_f \omega_f z_f(t) - 2\omega_f \dot{y_f}(t) - \omega_f^2 y_f(t) = 0 \quad \textbf{(3.14d)}$$

Chapter 4 : Audio-Visual Multisensory Integration using NMM

Modelling the audio-visual multisensory integration using NMM is the focus of this chapter and this thesis. The approach consists of Wendling's model but with self-loop in fast inhibitory interneurons ($GABA_{A,fast}$) as in (Ursino, Cona and Zavaglia, 2010). As summarized in chapter 1, the model replicated neuronal physiology as the individual sensory modalities are processed in their respective areas/regions in brain (unisensory processing) and then the information is fed to a higher hierarchical level that performs multisensory processing. In this case, the multisensory processing is considered to be done by multisensory neurons which are assumed at a higher hierarchical level.

4.1. A single ROI

Extending the previous work from Jansen and Rit model, and Wendling's model the current approach to generate alpha-gamma rhythms makes use of four interconnected ROIs (each with four neuronal populations: pyramidal, $GABA_{A,slow}$, $GABA_{A,fast}$, and excitatory interneurons) along with self-loops in fast inhibitory interneurons with each ROI receiving excitatory input (from pyramidal cells of other ROIs or any other external input) at pyramidal neurons and inhibitory input at fast inhibitory interneurons. Although the neurophysiology evidences that the input to a certain ROI or a cortical column can be received at both types of interneurons as well as pyramidal cells (David, Harrison and Friston, 2005), the study by (Ursino, Cona and Zavaglia, 2010) proved that only the inputs at $GABA_{A,fast}$ and pyramidal neurons are enough to generate oscillations in a ROI.

The model of a single ROI is represented in figure 4.1. The weights are denoted by C_{ij} wherein the first subscript is for post-synaptic population and the second subscript is for pre-synaptic population. Subscripts e, p, s, and f are respectively for excitatory interneurons, pyramidal neurons, $GABA_{A,slow}$, and $GABA_{A,fast}$ interneurons. Furthermore, the synaptic dynamics of excitatory interneurons, pyramidal neurons, and excitatory inputs are considered to be same ($G_e = G_p = G_I$, $\omega_e = \omega_p = \omega_I$). It can be seen that there are eight weight parameters (C_{ij}) in one ROI.

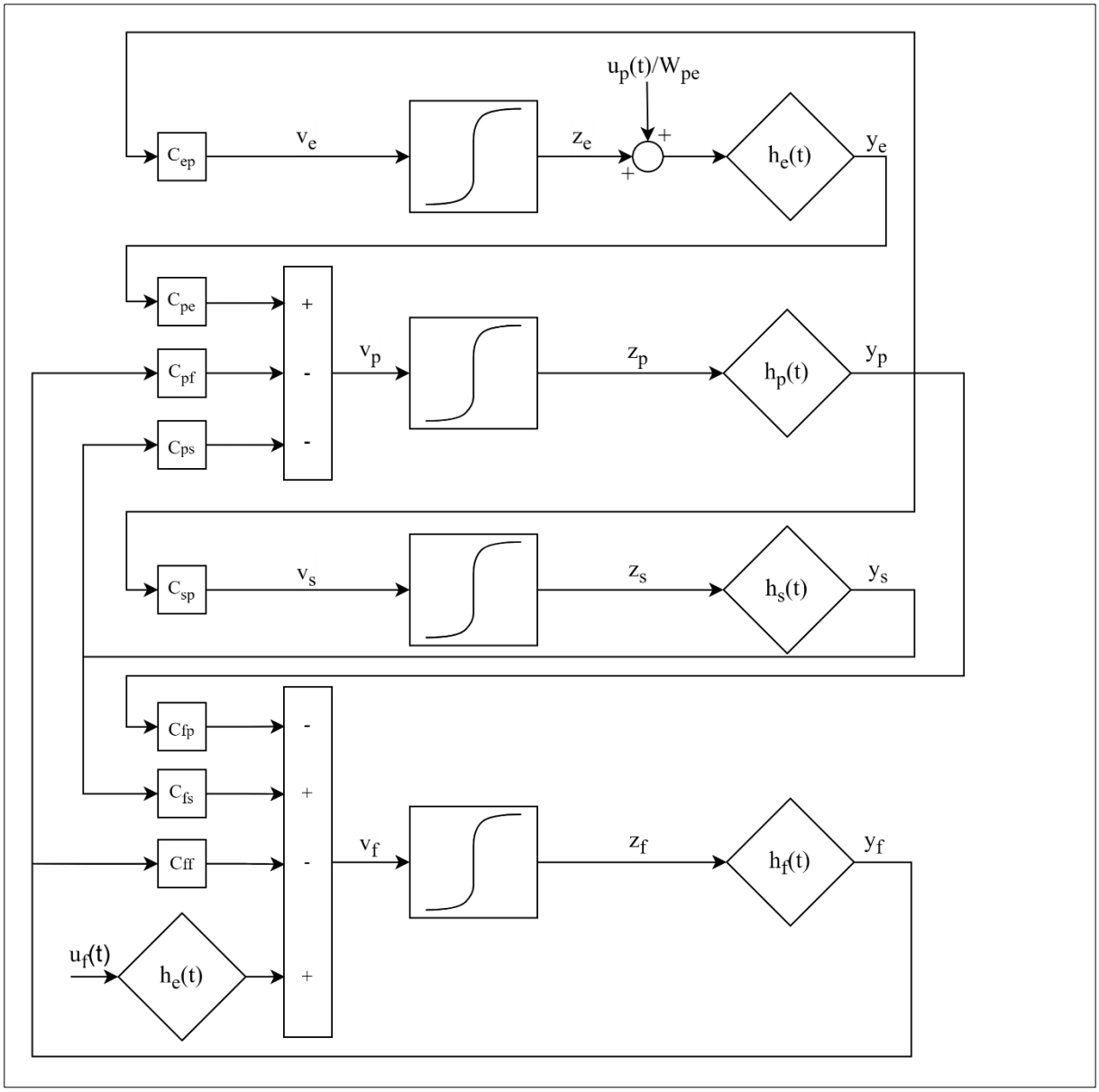

Fig.4.1. Single ROI of 4 four interconnected populations.

Now this model can be mathematically written as,

Pyramidal Neurons

$$v_p(t) = C_{pe}y_e(t) - C_{ps}y_s(t) - C_{pf}y_f(t) \quad \textbf{(4.1a)}$$

$$z_p(t) = S\left(v_p(t)\right) = \frac{2e_0}{1 + e^{-r(v_p(t) - v_0)}} - e_0 \quad \textbf{(4.1b)}$$

$$\dot{y_p}(t) = x_p(t) \quad \textbf{(4.1c)}$$

$$\dot{x_p}(t) = G_p\omega_p z_p(t) - 2\omega_p x_p(t) - \omega_p^2 y_p(t) \quad \textbf{(4.1d)}$$

Excitatory Interneurons

$$v_e(t) = C_{ep} y_p(t) \quad \textbf{(4.2a)}$$

$$z_e(t) = S\big(v_e(t)\big) = \frac{2e_0}{1 + e^{-r(v_e(t) - v_0)}} - e_0 \quad \textbf{(4.2b)}$$

$$\dot{y_e}(t) = x_e(t) \quad \textbf{(4.2c)}$$

$$\dot{x_e}(t) = G_e \omega_e \left(z_e(t) + \frac{u_p(t)}{C_{pe}} \right) - 2\omega_e x_e(t) - \omega_e^2 y_e(t) \quad \textbf{(4.2d)}$$

Slow Inhibitory Interneurons

$$v_s(t) = C_{sp} y_p(t) \quad \textbf{(4.3a)}$$

$$z_s(t) = S\big(v_s(t)\big) = \frac{2e_0}{1 + e^{-r(v_s(t) - v_0)}} - e_0 \quad \textbf{(4.3b)}$$

$$\dot{y_s}(t) = x_s(t) \quad \textbf{(4.3c)}$$

$$\dot{x_s}(t) = G_s \omega_s z_s(t) - 2\omega_s x_s(t) - \omega_s^2 y_s(t) \quad \textbf{(4.3d)}$$

Fast Inhibitory Interneurons

$$v_f(t) = C_{fp} y_{p(t)} - C_{fs} y_s(t) - C_{ff} y_f(t) + y_l(t) \quad \textbf{(4.4a)}$$

$$z_f(t) = S\left(v_f(t)\right) = \frac{2e_0}{1 + e^{-r(v_f(t) - v_0)}} - e_0 \quad \textbf{(4.4b)}$$

$$\dot{y_f}(t) = x_f(t) \quad \textbf{(4.4c)}$$

$$\dot{x_f}(t) = G_f \omega_f z_f(t) - 2\omega_f x_f(t) - \omega_f^2 y_f(t) \quad \textbf{(4.4d)}$$

$$\dot{y_l}(t) = x_l(t) \quad \textbf{(4.4e)}$$

$$\dot{x_l}(t) = G_l \omega_l u_f(t) - 2\omega_l x_l(t) - \omega_l^2 y_l(t) \quad \textbf{(4.4f)}$$

This "complete" model of a single ROI is now capable of representing a neural circuit. Our model shall consist of 4 ROIs: two with intrinsic gamma band denoting conscious sensory processing, one with intrinsic alpha band which modulates attention, and one ROI with both bands to simulate multisensory processing. The generation of either alpha or gamma rhythm by any ROI depends on the choice of parameters used for that ROI. Specifically, the time constants (and some synaptic weights too) are crucial in maintaining a rhythm. From table 4.1, the time constants of excitatory and fast inhibitory interneurons are kept low to generate gamma rhythm while higher values are used for alpha rhythm. It means that dynamics of both types of neurons need to be fast to generate faster rhythms. Conversely, the slow inhibitory interneurons' dynamics are slow in gamma band as compared to alpha band.

Moreover, the synaptic weightage from both slow interneurons and pyramidal neurons towards fast interneurons (C_{fs} and C_{fp}) are kept at higher value to generate gamma rhythm meaning, while weightage from slow interneurons to pyramidal neurons (C_{ps}) is kept low

when generating gamma rhythm. Apart from these, all other values are kept similar. These points identify the importance of synaptic connectivity among different neural populations that lead to generation of rhythms.

	Parameter	Value for Alpha	Value for Gamma
Gain	G_e	5.17	5.17
	G_s	4.45	4.45
	G_f	57.1	57.1
Reciprocal of Time constant (s^{-1})	ω_e	66	125
	ω_s	42	30
	ω_f	300	400

Table 4.1. Gain and time constants.

	Parameter	Value for Alpha	Value for Gamma
Synaptic Weights	C_{ep}	54	54
	C_{pe}	54	54
	C_{sp}	54	54
	C_{ps}	450	67.5
	C_{fs}	10	27
	C_{fp}	35	108
	C_{pf}	300	300
	C_{ff}	10	10

Table 4.2. Synaptic Weights.

	Parameter	Value
Sigmoidal Characteristics	e_0	2.5
	$r\ (mV^{-1})$	0.56
	v_0	15

Table 4.3. Sigmoidal Characteristics.

These values can be modified to even simulate the oscillatory behavior in theta, delta, or beta bands. Moving on with this model, a network of ROIs can now be modelled to simulate a neurophysiological phenomenon.

4.2. Model of Audio-Visual Multisensory Integration

Based on the previous summary, the current approach is to have four brain areas, two for unisensory processing, one for modulating attention, and one multisensory area. Figure 4.2 shows how these areas relate to each other.

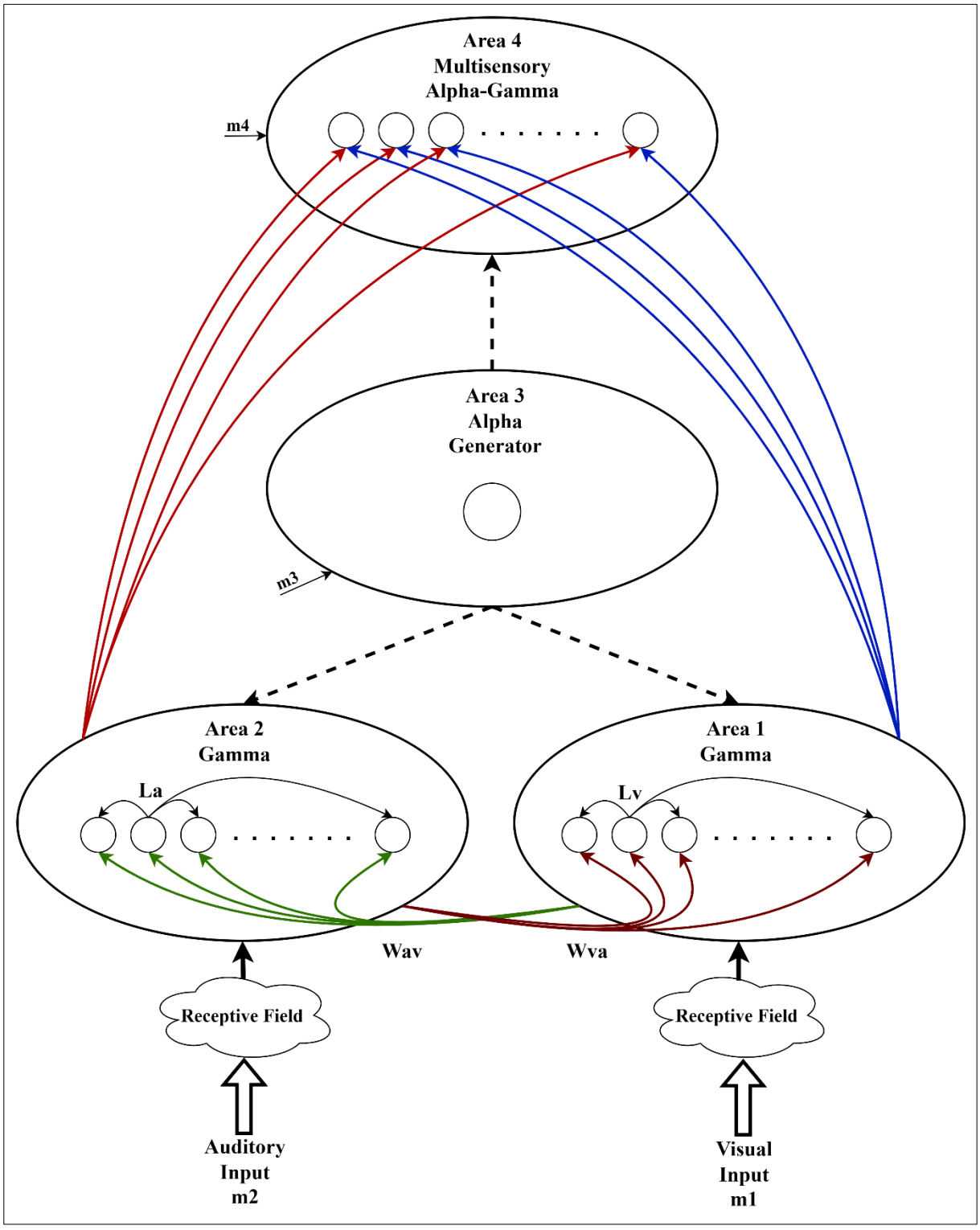

Fig.4.2. The complete model of 4 ROIs for audio-visual multisensory integration.

The inputs at unisensory areas are first passed through the receptive fields and are then processed by neural populations. These receptive fields and cross modal connections (bold black lines between Area 1 and Area 2) are trained (from a different model; details in subsequent topic) based on physiological definitions of auditory and visual information processing. The next level in hierarchy is a multisensory area which receives processed inputs from both unisensory areas and is able to simulate the physiology of ventriloquism, causal inference, and attention mechanism. These details will be further elaborated later. These areas receive further connection from another brain area that serves as the alpha rhythm generator to achieve attention mechanism.

A1 (Area 1) and A2 (Area 2) are unisensory areas corresponding to visual and auditory modalities, respectively, while A4 (Area 4) is the multisensory level. A3 (Area 3) generates alpha rhythms and is the basis of attention mechanism: It sends alpha rhythm to all other ROIs to enforce attention. Each area consists of 180 ROIs which account for 180^0 of spatial field in azimuthal direction, except for A3 which only has one ROI since it would send the same output (alpha rhythm) to every other ROI of all areas. Having this complete model, the inputs to each ROI are modelled by white gaussian noise plus the contribution of lateral and cross modal connections. Now obviously these connections are only valid for A1 and A2, while for A4 there are input connections coming from both unisensory areas and A3.

The equations 4.1-4.4 are replicated for all 180 ROIs in each area therefore matrix notations are used: Simple capital letters denote 180x180 matrices, bold capital letters represent 180xN arrays (***V*****$_i$**, ***Z*****$_i$**, ***Y*****$_i$**, ***X*****$_i$**, where *i = p,e,s,f*), small bold letters (***m1, m2,*** $\boldsymbol{w_{13}^f}$, $\boldsymbol{w_{23}^f}$, $\boldsymbol{u_{jk}}$, $\boldsymbol{n_{jk}}$, *where j=p,f and k=1-4*) represent 180x1 vectors, and simple small letters are scalar values. Note that there are two exceptions in these notations: zp3 and zf3, which are 1xN vectors, where N denotes the total number of discrete time samples. And since the variables are discretized in time, each variable is only taken at that instant of time, thus for example the variable $\boldsymbol{Z_{p2}}(t-\tau 1)$ is a 180xN array in which each column represents activity of all 180 neurons at a given sampled instant. The inputs are connected in a way that they can account for time delay within these connections and thus are modelled as,

$$\boldsymbol{u_{p1}}(\boldsymbol{t}) = \boldsymbol{n_{p1}}(\boldsymbol{t}) + L_{ex-v}\boldsymbol{Z_{p1}}(t-\tau 1) + W_{va}^{p}\boldsymbol{Z_{p2}}(t-\tau 1) + \boldsymbol{m1} \quad \textbf{(4.5a)}$$

$$\boldsymbol{u_{f1}}(\boldsymbol{t}) = \boldsymbol{n_{f1}}(\boldsymbol{t}) + L_{in-v}\boldsymbol{Z_{f1}}(t-\tau 1) + W_{va}^{f}\boldsymbol{Z_{f2}}(t-\tau 1) + \boldsymbol{w_{13}^f} z_{p3}(t-\tau 3) \quad \textbf{(4.5b)}$$

$$\boldsymbol{u_{p2}(t) = n_{p2}(t) + L_{ex-a}Z_{p2}(t-\tau 2) + W_{av}^{p}Z_{p2}(t-\tau 2) + m2} \quad \textbf{(4.6a)}$$

$$\boldsymbol{u_{f2}(t) = n_{f2}(t) + L_{in-a}Z_{p2}(t-\tau 2) + W_{av}^{f}Z_{f1}(t-\tau 2) + w_{23}^{f}z_{p3}(t-\tau 3)} \quad \textbf{(4.6b)}$$

$$\boldsymbol{u_{p3}(t) = n_{p3}(t) + m3} \quad \textbf{(4.7a)}$$

$$\boldsymbol{u_{f3}(t) = n_{f3}(t)} \quad \textbf{(4.7b)}$$

$$\boldsymbol{u_{p4}(t) = n_{p4}(t) + w_{41}^{p}Z_{p1}(t-\tau 4) + w_{42}^{p}Z_{p2}(t-\tau 4)} \quad \textbf{(4.8a)}$$

$$\boldsymbol{u_{f4}(t) = n_{f4}(t) + w_{43}^{f}z_{p3}(t-\tau 4)} \quad \textbf{(4.8b)}$$

Inputs *m1* and *m2* are spatial impulses at desired spatial positions (but constant in time) filtered through respective receptive field while *m3* is a constant value. The inputs to each ROI are delayed by some time lag, as can be seen in each population. However, note that $\tau 4$ is not a true delay but for sake of implementation a negligible value was used. Furthermore, the white gaussian noise inputs ($n_{ij}(t) \;\; i\epsilon\{p,f\}, j\epsilon\{1-4\}$) were chosen with zero mean and a variance of 5 units (μ=0, σ_i^2=5: $i\epsilon\{p,f\}$). It can be noticed that each input has the contribution from lateral synapses within an area and cross modal synapses between areas.

The variables $\boldsymbol{w_{13}^{f}}$ and $\boldsymbol{w_{23}^{f}}$ are related to modulating attention since they are cross-modal synapses from A3 to A1 and A2, respectively. A3 sends alpha rhythm to inhibit the influence of certain ROIs while focusing attention on others. Therefore, $\boldsymbol{w_{13}^{f}}$ and $\boldsymbol{w_{23}^{f}}$ are column vectors with 180 elements consisting of zeros which correspond to respective spatial positions. Wherever an inhibition is desired, we simply keep the corresponding elements equal to 45. So, for example if we need to inhibit the left half space (1-90^0) we can put the first 90 elements equal to 45 in both vectors. Conversely, if we want to inhibit one modality entirely while focusing attention on the other then one of the vectors (say $\boldsymbol{w_{13}^{f}}$ for inhibiting visual modality) needs to have all elements equal to 45 and the other vector with all zeros. In this way the desired attention mechanism is achieved.

Regarding the inputs, only two populations for every ROI are considered *pyramidal* and $GABA_{A,fast}$, which can receive connections from pyramidal neurons in other ROIs. Infact, the literature suggests that only pyramidal neurons send long range connections to other types of neurons in other ROIs. Now although the "other types of neurons" could be anyone among pyramidal neurons, $GABA_{A,fast}$, $GABA_{Aslowt}$ or excitatory interneurons (Felleman and Van Essen, 1991), (Ursino, Cona and Zavaglia, 2010) showed that only the first two types of inputs contribute significantly to the overall model dynamics. Thus, in our model, the

pyramidal neurons receive white gaussian noise, the spatial impulse filtered by respective RF, and the lateral and cross-modal contributions as in equations 4.5a and 4.6a, while no contribution from A3 i.e., the alpha rhythm. Whereas, from equations 4.5b and 4.6b we see that $GABA_{A,fast}$ neurons receive cross-modal connections from A3 and in turn inhibit the pyramidal neurons of that ROI. The reason for this type of connectionism is that $GABA_{A,fast}$ neurons have very fast dynamics which is why when they receive the cross-modal inputs from A3 they tend to preserve it and propagate to other neuronal populations (in this case to pyramidal neurons).

The arrays *Lin-a, Lin-v, Lex-a,* and *Lex-v* denote the lateral synapses within auditory and visual ROIs for both $GABA_{A,fast}$ and pyramidal populations respectively . These lateral connections are modelled to have gaussian characteristics as presented in equation 4.9.

$$L_{ex-J} = L_{ex0-J} * e^{-\frac{D(\theta_k,\theta)^2}{2\sigma_{ex}^2}} \qquad J\epsilon\{V,A\} \tag{4.9a}$$

$$L_{in-J} = L_{in0-J} * e^{-\frac{D(\theta_k,\theta)^2}{2\sigma_{in}^2}} \qquad J\epsilon\{V,A\} \tag{4.9b}$$

Here the terms $L_{ex0\text{-}J}$ and $L_{in0\text{-}J}$ represent strengths of lateral synapses, σ_{ex}^2 and σ_{in}^2 are the variances of excitatory and inhibitory connections, respectively. These connections realize a gaussian disposition. The notable point in the above equations is the use of circular connectivity, that is, neurons at position 1 and 180 are connected in a circular fashion so they excite (or inhibit) other neurons in a gaussian manner. For example, the neurons encoding position 1 are at the same distance from neurons at positions 2 and 180, or from neurons at positions 3 and 179. Equation 4.10 shows the model of such connectivity.

$$D_k(\theta_k,\theta) = \begin{cases} |\theta_k - \theta| & if|\theta_k - \theta| \leq 90 \\ 180 - |\theta_k - \theta| & if|\theta_k - \theta| > 90 \end{cases} \tag{4.10}$$

Here θ_k is the current neuron position while θ spans from 1 to 180 for each *k (k=1…180)*. Note that the quantity $D(\theta_k,\theta)$ is a matrix of 180x180 and $D_k(\theta_k,\theta)$ is the kth column of matrix $D(\theta_k,\theta)$. Therefore, the arrays L_{ex-J} and L_{in-J} are also matrices of same shape. A summary of input values is given in table 4.4.

Parameter	Value	Parameter	Value
m1	*intensity_v**800**Rec*$_v$****m1***	w_{43}^f	200
m2	*intensity_a**800**Rec*$_a$****m2***	w_{41}^p, w_{42}^p	300
W_{av}^f	7*W_{av}	L_{ex0V}, L_{ex0A}	12

W_{va}^{f}	7*W_{va}	L_{in0V}, L_{in0A}	9
$\tau 1$	20ms	$\tau 3$	50ms
$\tau 2$	20ms	$\tau 4$	0.1ms
σ_{ex}^{2}	2.5*2.5	σ_{in}^{2}	9*9
m3	1000	*intensity_v*	I*|pos_v-90| + 0.3
I	1/90/2	*intensity_a*	0.7*δ*(n-pos_a)*

Table 4.4. Input Parameters. Equations 4.5 – 4.9.

The right-hand members in first two rows (***m1*** and ***m2*)** are simply spatial impulses at desired positions and then filtered by RF to produce the right-hand members and thus the notation has been kept the same. Note that in these inputs the terms *pos_v* and *pos_a* are simply the positions between 1^0 and 180^0 (inclusive) that we want to excite, and in the value of *intensity_a*, the small delta represents an impulse at position *pos_a* where n is the sample number of ***m2***. Parameters W_{va}^{k} and W_{av}^{k} (where *k=p,f*) are cross-modal connections between A1 and A2, w_{43}^{f}, w_{41}^{p} and w_{42}^{p} are cross modal connections between different areas with first subscript denoting post-synaptic area and second subscript refers to pre-synaptic area, Rec_a and Rec_v are receptive fields of auditory and visual areas, respectively. All these input values are taken from (Ursino, Cona and Zavaglia, 2010), except that some modifications are made to adapt the inputs to the characteristics of the neural mass model.

These modifications include the synaptic time delays in A1 and A2 (denoted as $\tau 1$ and $\tau 2$, respectively) which are kept at 20ms, whereas the delay $\tau 3$ from A3 is kept at half of alpha period i.e., 50ms. Through multiple simulations it was observed that longer delays (in A1 and A2) lead to slower rhythm generation by either area. Then the lateral connections' variances and amplitudes were varied. At first, we selected same values as in (Ursino, Cona and Zavaglia, 2010) but the oscillations were not consistent with our desired goals thus the we increased the amplitudes and inhibitory variance while decreased the excitatory connections' variance. We found that the model needed inhibitory activations covering large number of ROIs but with smaller amplitude compared to excitatory connections that needed to be higher in amplitude but with shorter coverage to other ROIs. In this way a harmonic oscillatory behavior was achieved.

The receptive fields were weighted slightly differently since they are obtained from (Ursino *et al.*, 2017) while from the NMM model of (Ursino, Cona and Zavaglia, 2010) it was seen that the inputs had an amplitude of 800. Moreover, the inputs now are not simply spatial impulses

but rather filtered through RF and so a wide band of neurons are active at same time thus we tuned the weights to achieve optimal oscillatory behavior of our NMM. The cross-modal connections, too, were tuned in this manner. Cross-modal connections with low weights did not have any significant effect on the target area (such as change in auditory position during ventriloquism effect) while higher weights led to collapse of oscillations altogether. Hence, using different values, we reached the optimally weighted connections for the model. The model was first tested without cross-modal connections and once it worked well with chosen lateral connections' amplitude and variances only then the cross-modal inputs were also included and tuned accordingly.

Finally, the input of visual modality was modulated by a linear trend to ensure the peripheral positions receive higher intensity input compared to central positions. In fact, the receptive field's maximum amplitude at each position decreases from center to periphery and in order to get the system working in desired oscillatory condition it is necessary to increase the input at the periphery. Physiologically, this means that to excite the neurons near fovea a lighter intensity input would suffice while exciting peripheral neurons, comparably, a higher intensity input is required.

4.3 Training Receptive Fields and Cross-Modal Connections

The training was done in a previous work (Ursino *et al.*, 2017) using Hebbian learning with a decay term. These values are directly used in this model with just some weight factors to optimally simulate the model in desired conditions. For auditory receptive field (Rec_a) the authors started with inputs having random distribution superimposed with noise and with equal strengths. Thus, after training, the RFs for each of 180 positions were distributed similarly without any significant changes in amplitude or shape. In the case of visual receptive field (Rec_v) the initial setup consisted of having more visual inputs near fovea compared to the peripheries and thus after training the RFs near fovea (central position) were sharp and had high amplitude while more broader RFs but with less amplitude were evident in the peripheral regions. Moreover, the barycenter of the RFs shifted towards the fovea. This reflects the idea of having higher likelihood to encounter an input near fovea as compared to peripheries.

For cross-modal synapses, during training, there was a certain percentage of cross-modal stimuli, with visual and auditory inputs close in space. Moreover, the prior information of

preferred positions from visual and auditory stimuli was necessary to train them. Since the visual neurons had sharper RFs near fovea, the cross-modal connections from visual to auditory neurons after training resulted in higher excitation of auditory neurons closer to central region whereas the peripheral regions received less excitation. On the other hand, the cross-modal connections from auditory to visual neurons resulted in similar excitation profiles throughout most of the azimuthal space except at the peripheries which were highly excited. All the results are detailed in the next chapter.

4.4. Calculating Position from Oscillations

The final step for the model was to calculate the spatial position given the oscillations of some neurons. This step was calculated using the barycenter method. First, for each neuron the preferred spatial position was calculated using equation 4.11.

$$K_i^s = \frac{\sum_{j=1}^{180} r_{ij} * p_j}{\sum_{j=1}^{180} r_{ij}} \qquad s = A, V \tag{4.11}$$

Here r_{ij} is the receptive field of i^{th} neuron and p_j is the distance in degrees calculated as,

$$p_j = \begin{cases} [1\ to\ 89, j - 90\ to\ 0] & j < 90 \\ [181\ to\ j + 90, j - 89\ to\ 180] & j > 90 \\ [1\ to\ 180] & j = 90 \end{cases} \tag{4.12}$$

Then the preferred position was shifted to ensure the left (1 to 90 degrees) vs right (91 to 180 degrees) hemifields were equidistant from the central position (90th degre). To do so the prior input position information (*pos_s*, where *s=a,v*) was used.

$$q_j = K_j^s \tag{4.13a}$$

$$q_j = \begin{cases} \left[K_j^s > (pos\_s + 90)\right] - 180 & pos\_s < 90 \\ \left[K_j^s < (pos\_s - 90)\right] + 180 & pos\_s > 90 \\ K_j^s & pos\_s = 90 \end{cases} \tag{4.13b}$$

Equation 4.13b shows that the values in K_j^s which are greater than those in square brackets are subtracted (or added) by 180 to ensure the positions in left or right hemifields remain consistent. Finally, given the oscillations the barycenter, as perceived by brain, is calculated as,

$$B^s = \frac{\sum_{j=1}^{180} an_j^s * q_j}{\sum_{j=1}^{180} an_j^s} \qquad s = A, V \tag{4.14}$$

The meaning of an_j^s is “activity of neurons” which is the mean activity of neurons in a certain time window. For ROI1 and ROI2 (corresponding to visual and auditory modalities) the time window was selected to be 80ms because it works in gamma rhythm and for ROI4 the window is set at 200ms since it works in alpha rhythm. These intervals correspond to approximately two periods of the corresponding rhythms.

Chapter 5 : Simulation and Results

5.1. Receptive Fields and Cross-Modal Connections

Previous implementations of the model were coded in MATLAB software. Firstly, from the work of (Ursino, Cona and Zavaglia, 2010), plots of receptive fields and cross-modal connections are shown in figures 5.1 to 5.4. In figures 5.1 and 5.2 the first plots refer to receptive fields corresponding to respective neurons. They are plotted for ROI 10 to 180 with 20-unit steps (like 10,30,50...180). The second plots of each figure display the receptive fields of neuron number 80. It is evident that in visual RF the peak of the 80th neuron shifts to the position of 85, showing that the density of neurons is higher close to the fovea than in periphery. Another peculiarity is the shape of RFs: it can be seen that near the fovea the RFs are sharp with higher amplitudes compared to peripheries where the RFs are wider with lower amplitudes.

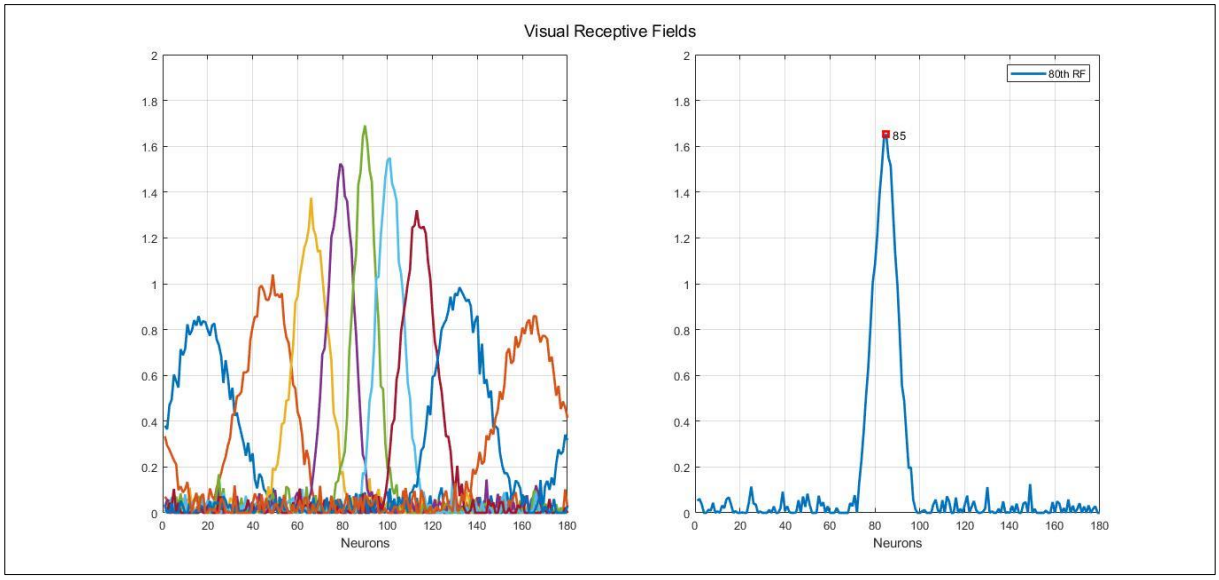

Fig. 5.1. Visual Receptive Fields.

Conversely, for the auditory modality the RFs are quite consistent in terms of shape and strengths and are uniformly distributed denoting the fact that auditory neurons encoding spatial positions are almost equidistant from each other. In figure 5.3 the preferred position of ROIs encoding for spatial positions are shown for both modalities. It is now clear from the first plot that the visual neurons tend to be biased towards the fovea or, in other words, there are more neurons encoding spatial positions closer to fovea compared to peripheral positions

where there are scarce neurons. In the case of auditory neurons, the receptive fields are quite linearly spaced from each other and thus encode spatial positions in an equidistant manner.

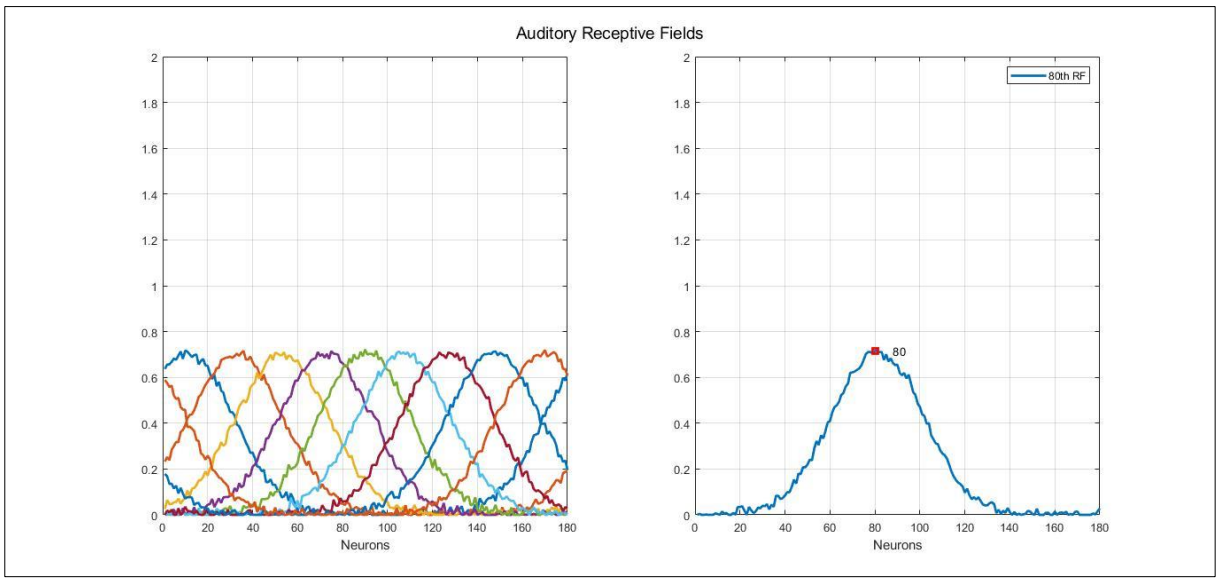

Fig. 5.2. Auditory Receptive Fields.

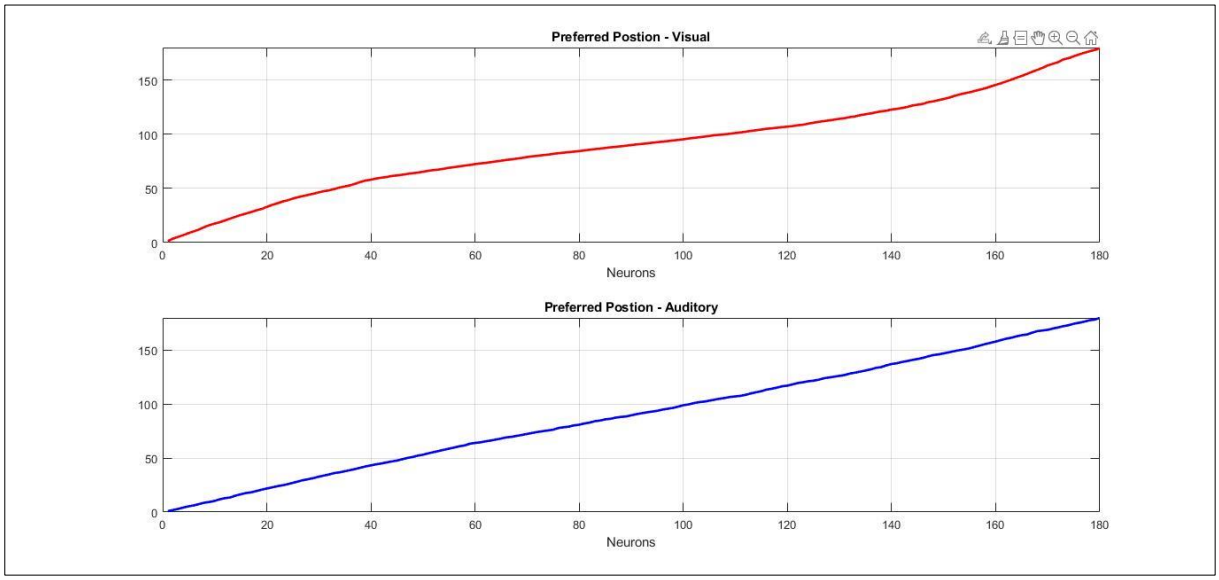

Fig. 5.3. Preferred spatial position of neurons.

Concerning cross-modal connections, as previously stated, the cross-modal connections from auditory to visual neurons excite visual neurons in a similar way except for the peripheral ones which they excite the most. The connections from visual neurons excite the central neurons in auditory modality more than the peripheral ones, as evident from figure 5.4. Note that these plots are plotted in a similar manner as in figures 5.1 and 5.2, that is, the connection from each neuron (to other modality) are plotted from 10^{th} to 180^{th} neuron with 20-unit steps.

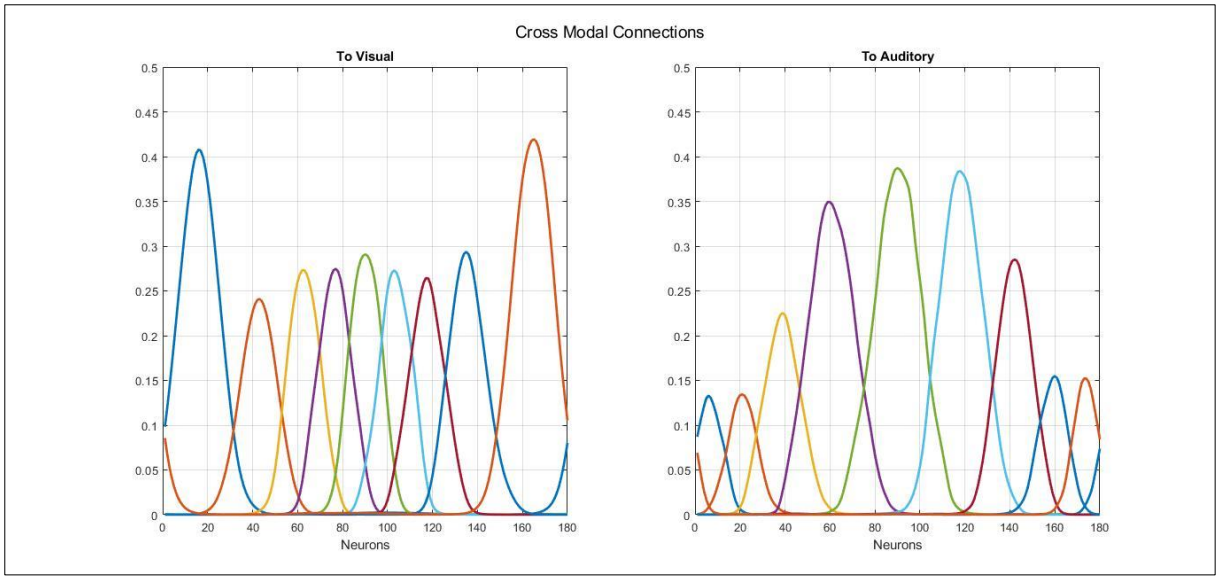

Fig. 5.4. Cross-Modal Connections.

5.2. Inputs

Inputs to both ROI1 and ROI2 are filtered through receptive fields. To start we give an impulse at desired spatial positions (70^0 for auditory and 110^0 for visual modality in this example as shown in figure 5.5). Each ROI then receives the input (which is 1x180 column vectors) scalar multiplied with its receptive field. Thus, the inputs reaching each ROI assume gaussian type shape, where the neurons close to spatial impulse are the ones that are most excited.

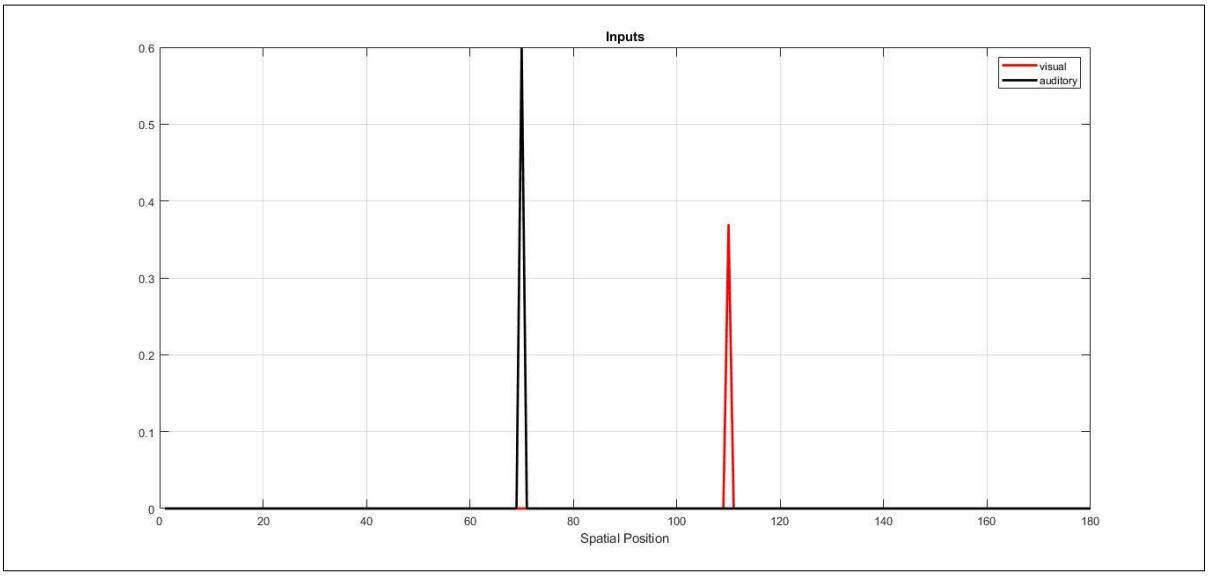

Fig. 5.5. Inputs.

Note that since each modality has different RFs, thus, to excite each ROI we need appropriate amplitude of input, therefore we see a lower value for visual modality and higher value for auditory modality.

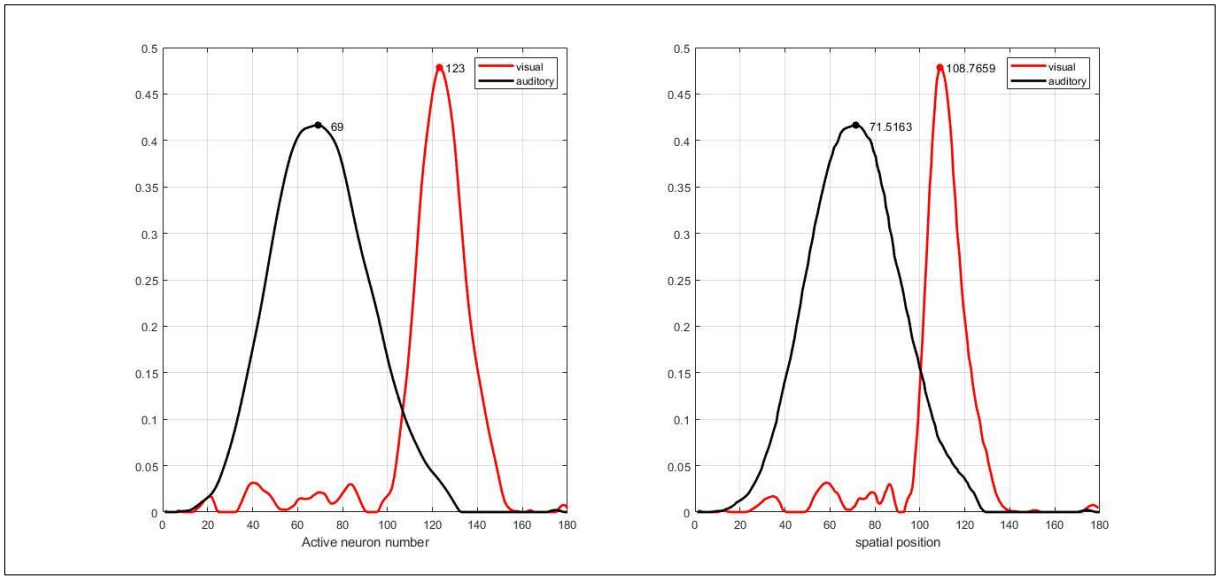

Fig. 5.6. Inputs processed through RFs.

Figure 5.6 depicts the picture of inputs after being filtered by RFs. The first plot simply represents what neurons are excited when a certain spatial position is excited. As discussed previously, the neuron density is higher closer to fovea compared to peripheries thus the neuron most excited by impulse at 110^0 is 123^{rd} neuron for visual modality while no appreciable difference is observed for auditory modality, consistent with what was said earlier. These inputs (filtered by RF) are represented with respect to spatial position in the second plot of figure 5.6 and we can see the peak values are close to 70 and 110 degrees, for auditory and visual modalities, respectively.

5.3. Oscillations and Position

For each ROI, the spike rate is the variable of interest for analysis (Note that the spike rate of pyramidal population is of concern and subsequent use of "spike rate" would imply the spike rate of pyramidal populations only). From previous discussions, it is now evident that the unisensory regions would work in gamma band reflecting the conscious processing while if any of them is inhibited, that is, less attention or focus is put on it by brain, then the activity becomes gamma modulated by alpha. This is the case for ROI1 and ROI2 while ROI3 is only an alpha rhythm generator thus working only in alpha band and ROI4 reflects the

multisensory processing, always working in gamma modulated by alpha rhythm. Figures 5.7 and 5.8 show the spike rates of visual and auditory modalities, respectively. In both figures the first plot depicts the activity of all 180 neurons while the second plot shows the most excited neuron's activity. Ignoring the initial transient period, the activity of neurons that are excited can be seen oscillating in phase but with slightly different amplitudes which are evident from the first plots of both figures.

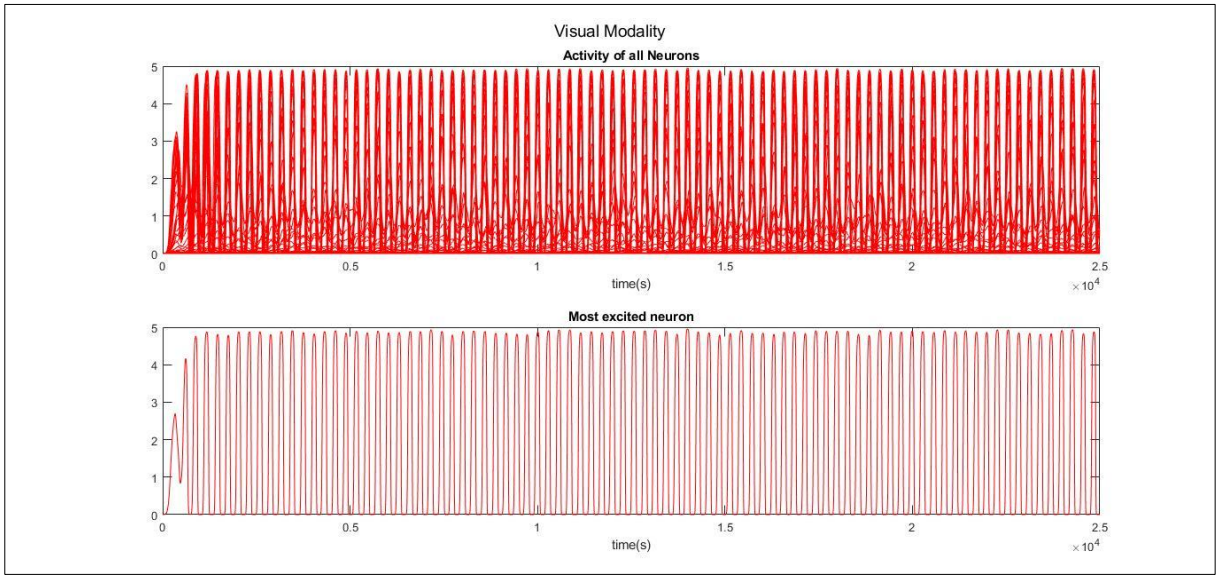

Fig. 5.7. Spike rate of visual modality.

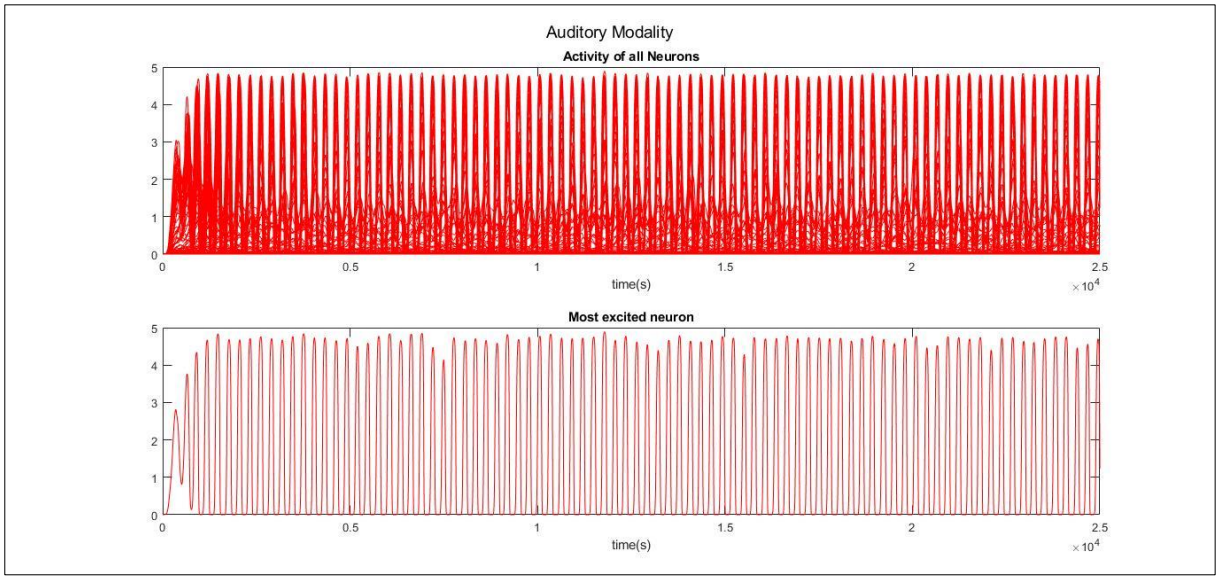

Fig. 5.8. Spike rate of auditory modality.

Moving on, the plots for ROI3 and ROI4 are given in figures 5.9 and 5.10. In 5.9, there is only one neuron population since its only role is to generate alpha rhythm while in 5.10 the first plot shows the activity of all multisensory neurons and the gamma modulated by alpha

rhythm can be seen. The second and third plots of 5.10 show the most excited neurons that come from both ROI1 and ROI2, since these both ROI excite the similar positioned neurons in multisensory region.

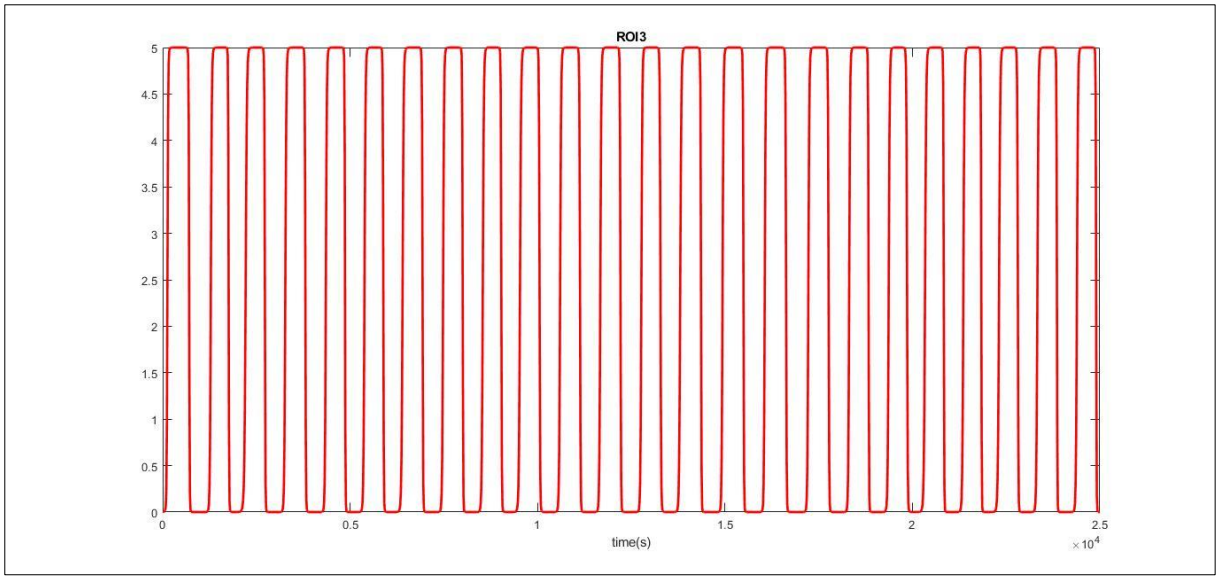

Fig. 5.9. Spike rate of ROI3 (Alpha rhythm generator).

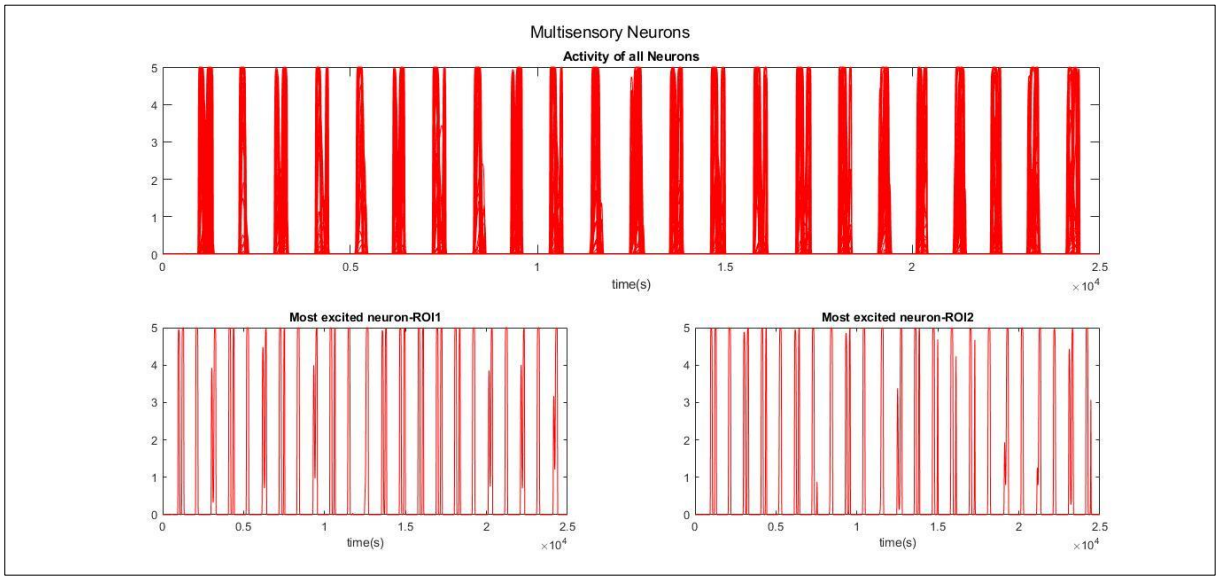

Fig. 5.10. Spike rate of multisensory neurons.

To get a better idea about frequency the PSD of these spike rates is shown in figure 5.11. Note that these signals were high pass filtered at 4 Hz passband to remove the initial transient effect. Both ROI1 and ROI2 are working in gamma band (around 35Hz), ROI3 is working around 10Hz as expected, and finally the multisensory ROI is oscillating in gamma modulated by alpha rhythm, which is evident from the two peaks, one at 10Hz and the other at around 35Hz. Particularly, the ROI4 seems to have two major peaks at 10Hz and 35Hz but there is

also a comparable peak at around 20Hz which could be due to the slight uneven interplay of ROI1 and ROI2 leading to this effect in ROI4. ROI4 provides the PSD of both barycenter 1 and 2 that respectively refer to the two unisensory inputs.

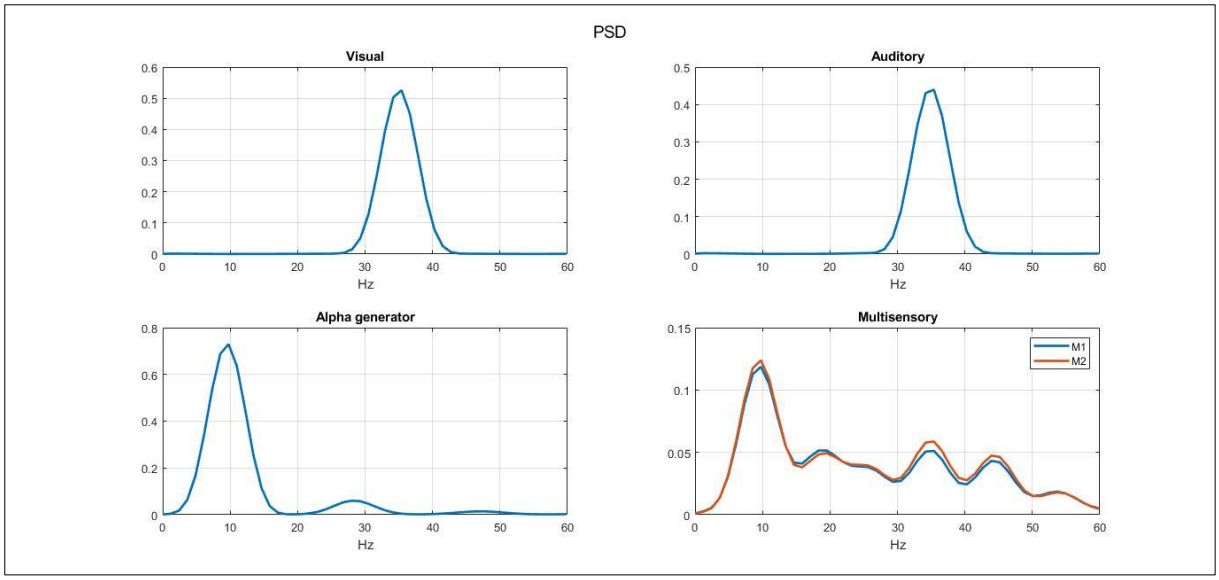

Fig. 5.11. PSD of spike rates of all ROIs.

From the oscillatory behavior, the position information was calculated as defined in topic 4.4 and plotted in figure 5.12.

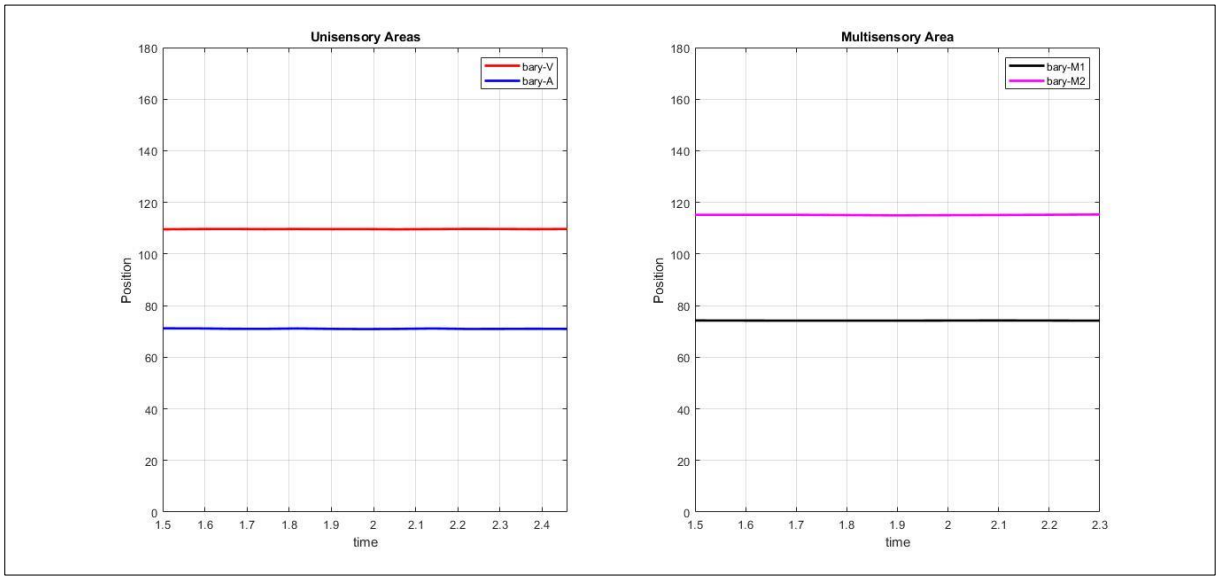

Fig. 5.12. Spatial position calculated from oscillations.

The barycenter reflects the position information of each ROI. From the first plot it is clear that given the inputs of auditory and visual modalities at 70^0 and 110^0, respectively, the barycenter at each time step in the two unisensory areas show almost the same position perception. The

barycenter is plotted considering the average of signal in a certain time window. For unisensory regions the barycenter is quite accurate but for the multisensory regions it is not the case. There is some error in calculating the position in multisensory area. One reason for it is that when calculating the barycenter in multisensory area, we simply took into account the average of both RFs of unisensory modalities. This approach is not quite physiological since multisensory neurons have a different spatial register for RFs of different modalities (further details in next chapter).

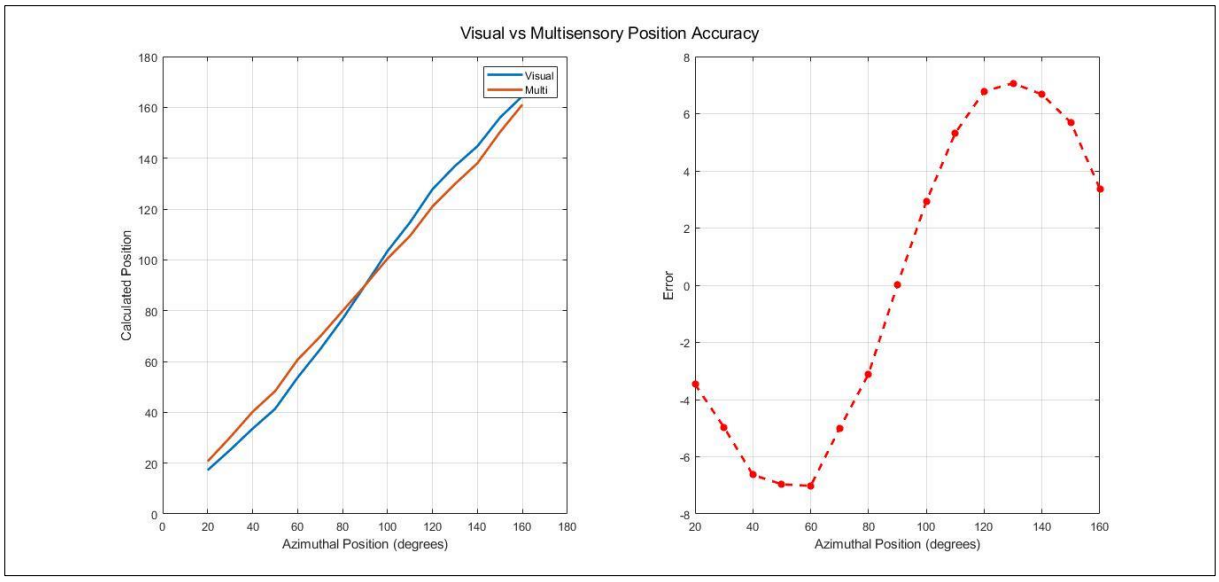

Fig. 5.13. Position accuracy in unisensory (visual) vs multisensory area.

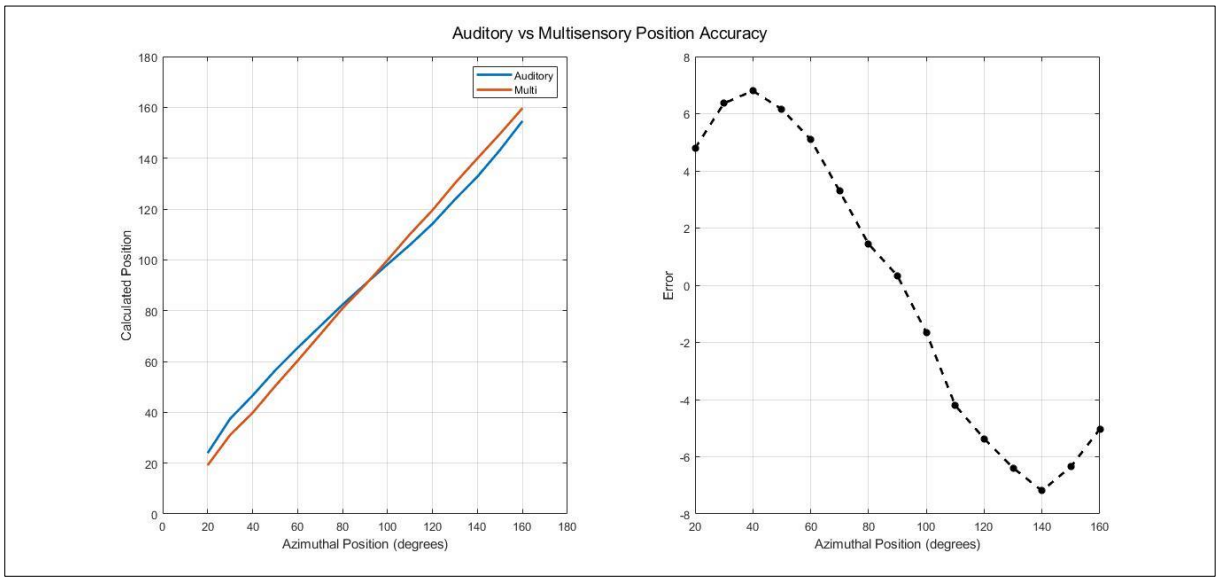

Fig. 5.14. Position accuracy in unisensory (auditory) vs multisensory area.

To better represent the position accuracy for each area we simulated the model for each unisensory input ranging from 20^0 to 160^0 (not included very extreme values since some

neurons at the opposite ends were excited due to use of circular distance which caused improper calculation of position information). Figures 5.13 and 5.14 represent the position information calculated for each unisensory modality separately and compared with multisensory position information. The errors in multisensory position perception, when a single unisensory input is applied, are plotted in the right-sided subplots in both figures.

5.4. Causal Inference

To get the model's behavior regarding causal inference the first attempt for analysis consists of keeping the visual input around fovea (at 90^0) and varying auditory input location to see when the model starts to perceive two inputs as one. The perception of causes is processed in multisensory layer and although there is a slight error in spatial localization in multisensory layer, nonetheless, the identification of cause(s) is evident. At about 53^0 (37^0 distance between A-V inputs) the differentiation between causes disappears in dual-modality setup. The results shown in figure 5.15 depict the situation in detail.

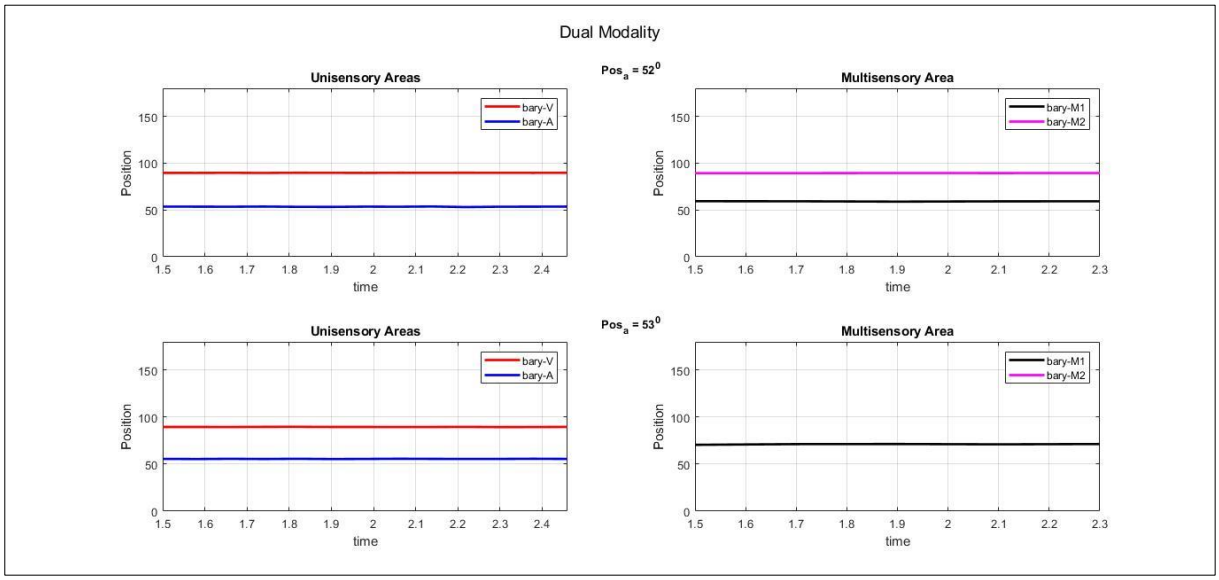

Fig. 5.15. Causal Inference (Pos_v = 90^0,Pos_a = [52^0, 53^0]).

In case of more peripheral inputs the situation changes: The ability to perceive separate causes increases. Figure 5.16 shows that the ability to differentiate causes in dual modality inputs stands at 24^0. Similar situation can be witnessed in the left periphery (1-90^0) since the neurons are connected in circular fashion with symmetric RF on left and right sides so for brevity the results are not shown here.

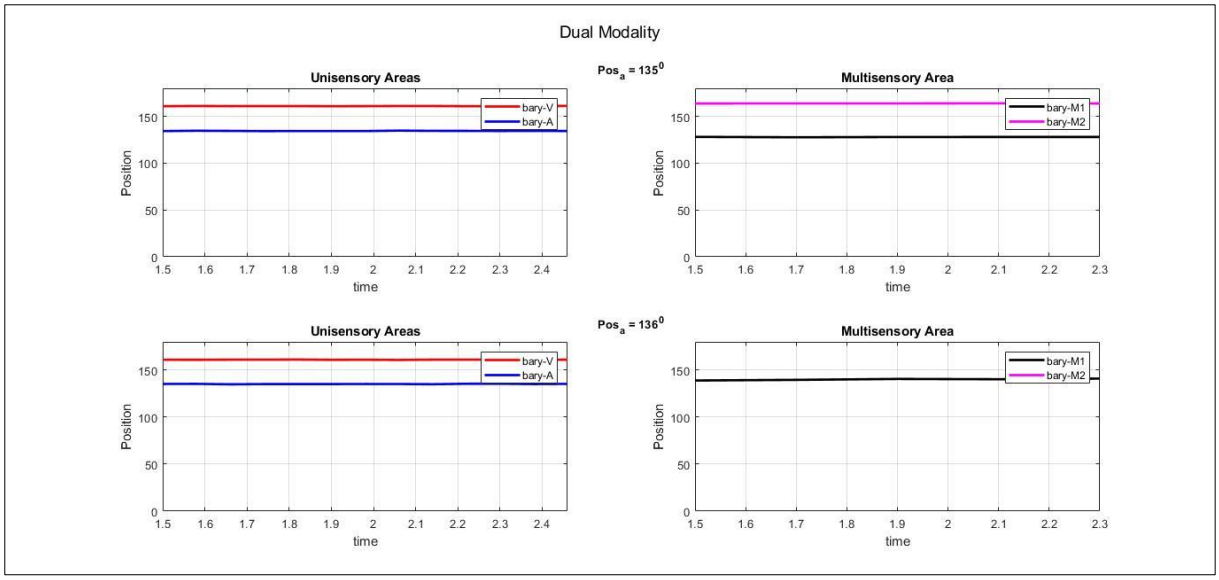

Fig. 5.16. Causal Inference (Pos_v = 160^0,Pos_a = [135^0,136^0]).

In case of inputs only in a single modality, both the unisensory and multisensory areas can provide causal description. For visual modality the ability to differentiate inputs around fovea stands at 17^0 as shown in figure 5.17. One visual input is at 90^0 and the other at 108^0, both the unisensory and multisensory areas can distinguish between them while there is no distinction between two inputs when one input is at same position (90^0) and the other input is applied at 107^0. This is evident in both unisensory and multisensory areas. On the other hand, when the inputs are in peripheries the situation changes. The ability to distinguish between inputs decreases to 37^0 as represented in figure 5.18.

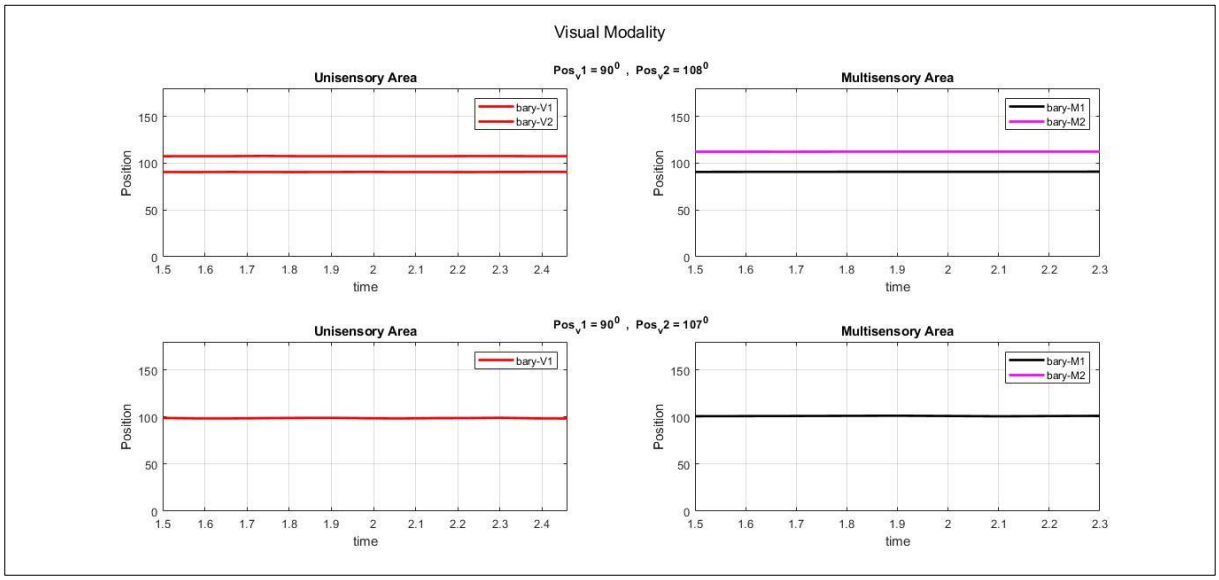

Fig. 5.17. Causal Inference in visual modality around fovea (Pos_v1 = 90^0, Pos_v2 = [108^0, 107^0]).

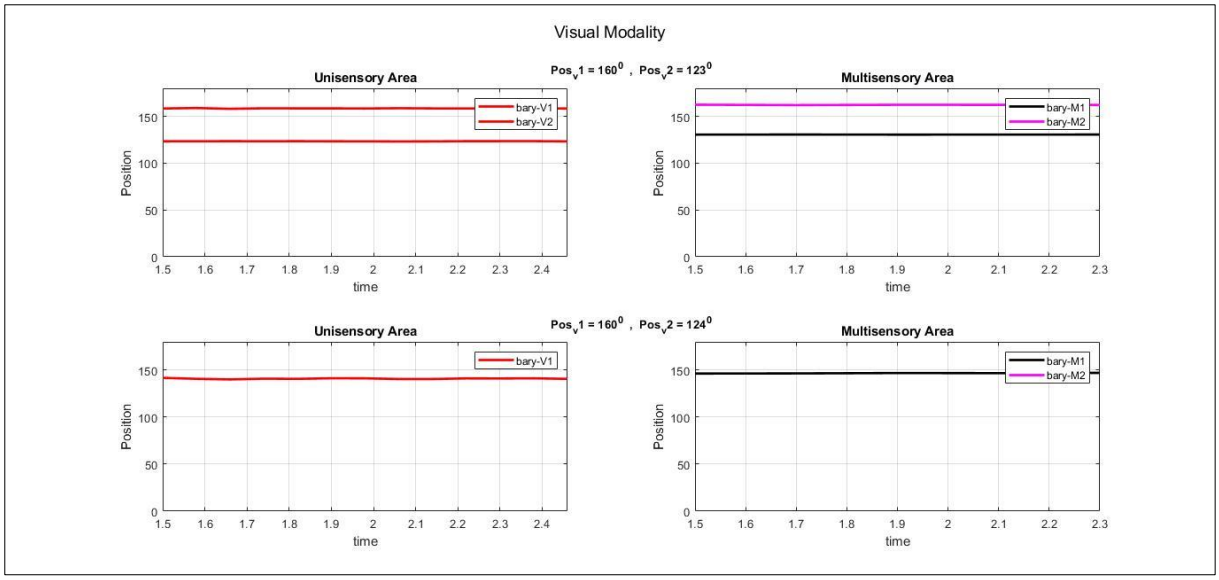

Fig. 5.18. Causal Inference in visual modality around periphery (Pos_v1 = 160^0, Pos_v2 = [123^0, 124^0]).

In case of auditory modality, since the RFs are uniform throughout the azimuthal space, there is no distinction between central or peripheral regions and thus only one example is shown in figure 5.19. The ability to distinguish between inputs stands at around 61^0-63^0. Note that we tried to simulate the same results by applying second auditory input (*Pos_a2*) at both 152^0 and 153^0 but there was no appreciable change in multisensory area, as can be seen in figures 5.20 and 5.21, thus we increased the range by a couple of degrees. Finally, it should be noted that the RFs are symmetric in the left and right side of azimuthal space so the same results can be obtained on the other side.

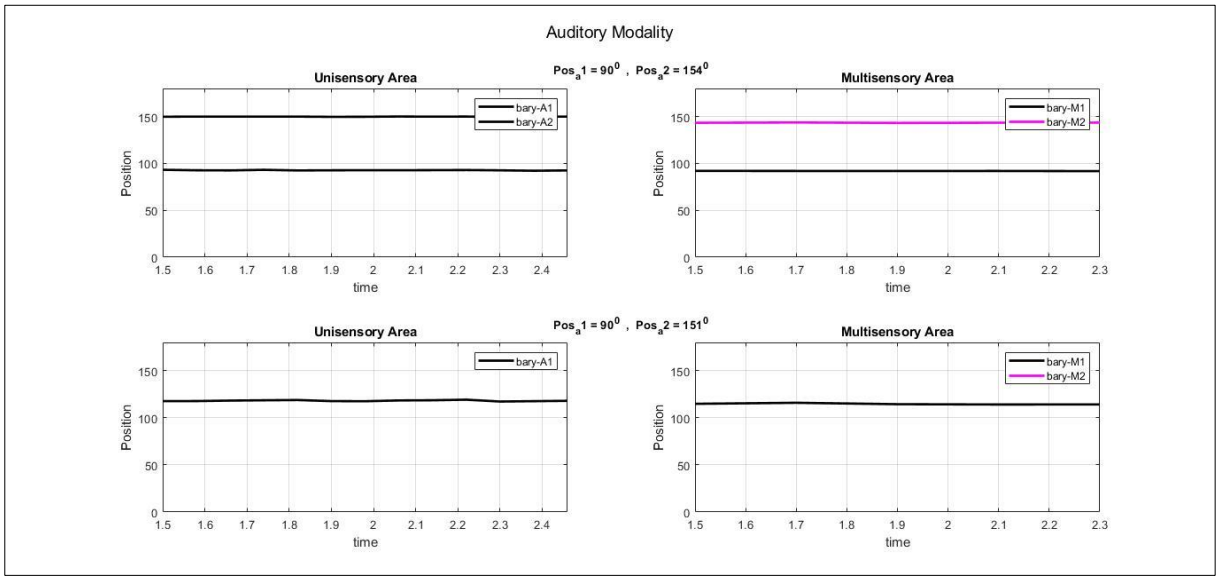

Fig. 5.19. Causal Inference in auditory modality (Pos_a1 = 90^0, Pos_a2 = [154^0, 151^0]).

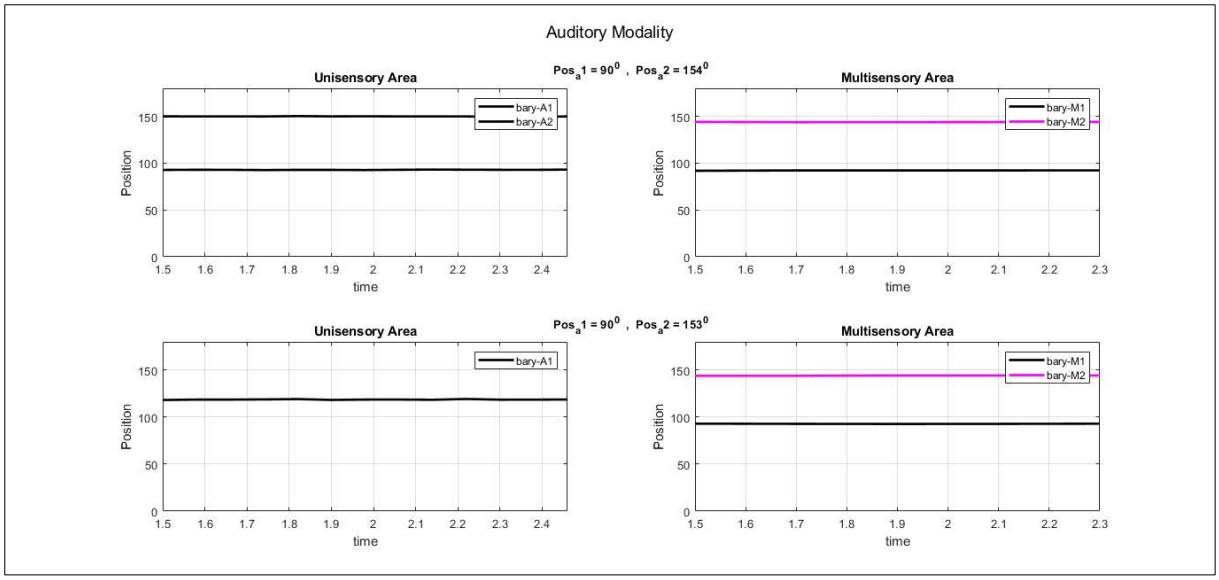

Fig. 5.20. Causal Inference in auditory modality (Pos_a1 = 90^0, Pos_a2 = [154^0, 153^0]).

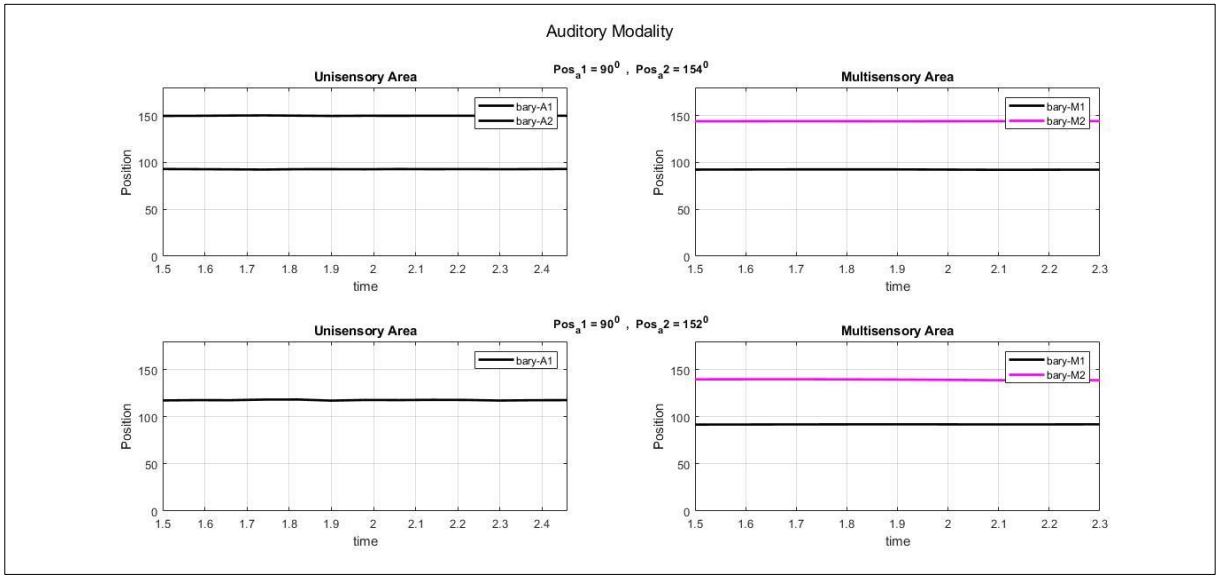

Fig. 5.21. Causal Inference in auditory modality (Pos_a1 = 90^0, Pos_a2 = [154^0, 152^0]).

5.5. Ventriloquism

To simulate ventriloquism effect, we put the visual input at 90^0 while the auditory input was varied from 50^0 to 130^0 with an increment of 10^0, and the position error at each instant was plotted against audio-visual position difference as shown in figure 5.22. The error in visual position remained less than 1^0 while the auditory position had significant error. This error was highest (4^0) at ± 20^0.

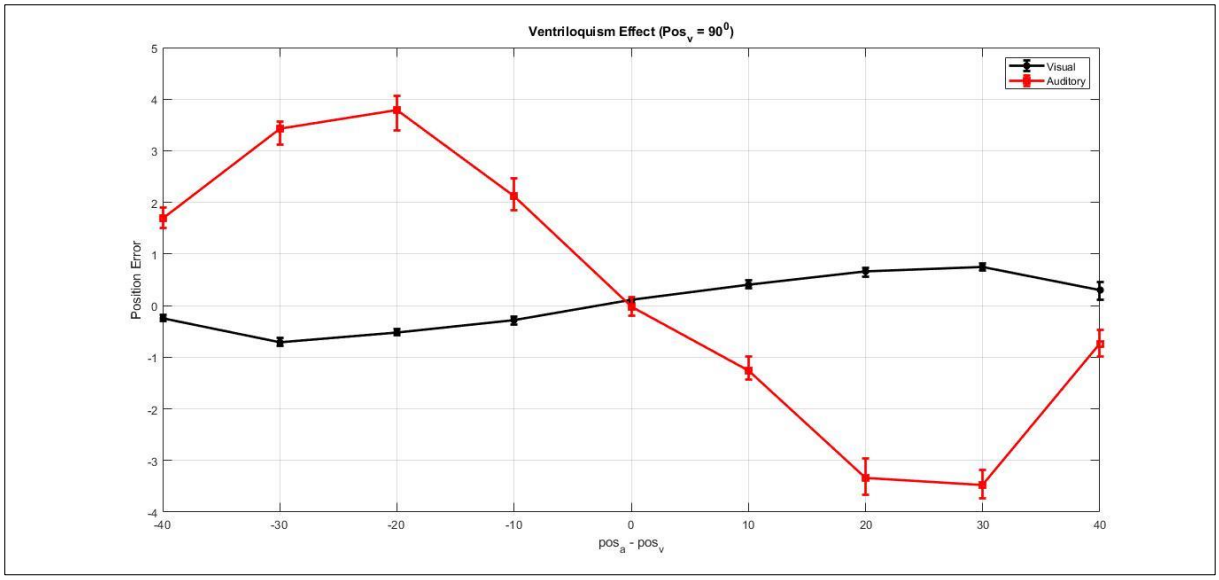

Fig. 5.22. Ventriloquism (Pos_v = 90^0, Pos_a = [50^0-130^0]).

5.6. Attention Modulation

In the first instance, we put all elements in column vector $\boldsymbol{w_{13}^f}$ equal to 0 and in $\boldsymbol{w_{23}^f}$ equal to 45. This ensures that the alpha rhythm is transmitted to A2 (auditory modality) only, resulting in inhibition of that area. The inputs for visual and auditory areas were put at 110^0 and 70^0, respectively. It can be seen from figure 5.23 that the unisensory areas process the respective inputs while in the multisensory area the auditory information is inhibited.

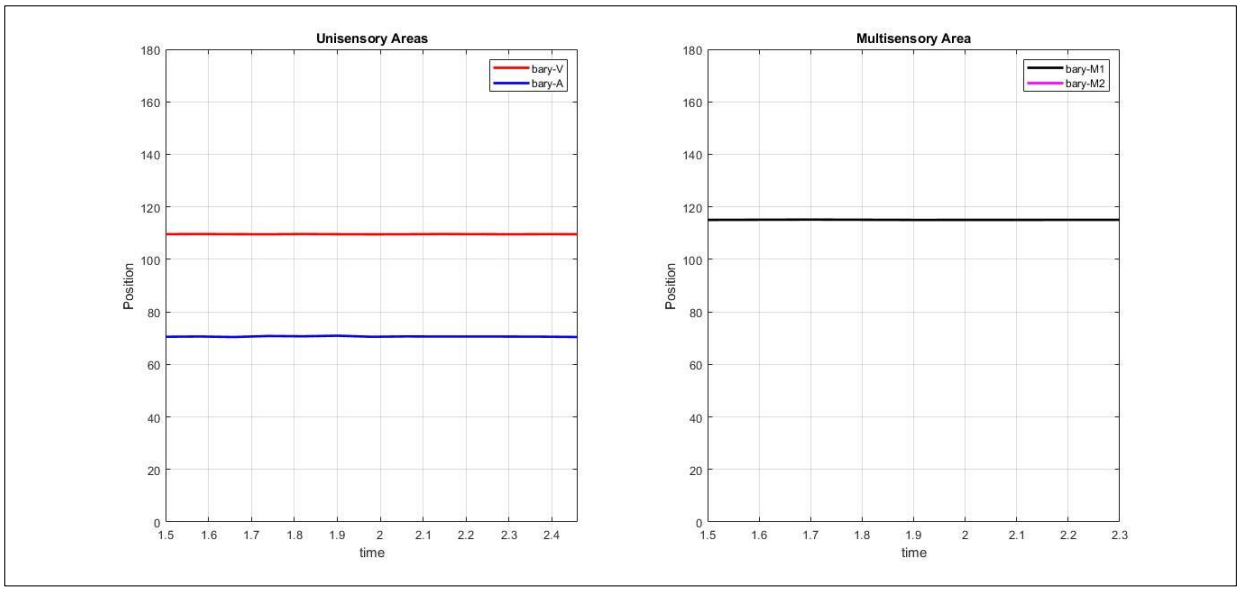

Fig. 5.23. Spatial position in unisensory and multisensory areas during attention modulation.

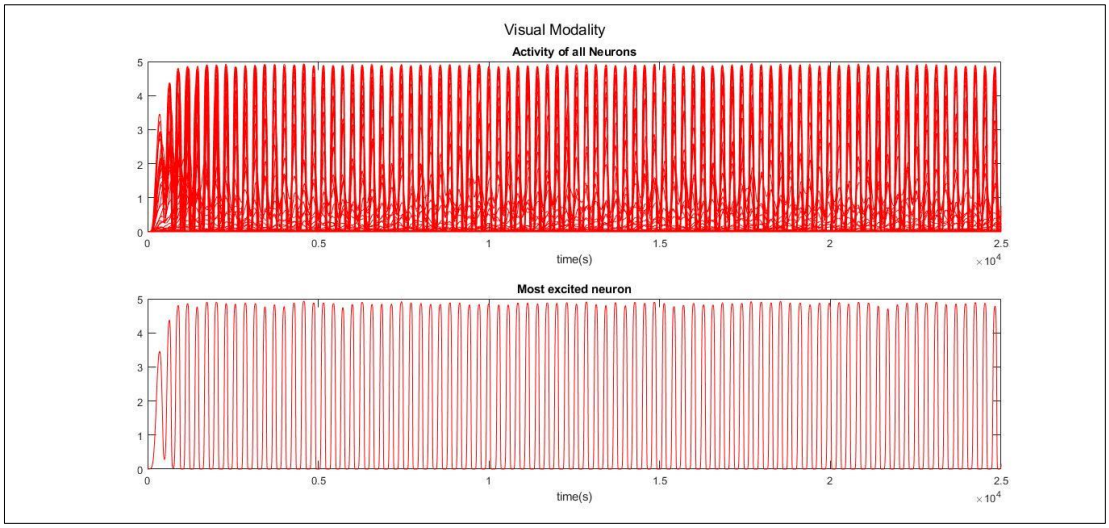

Fig. 5.24. Spike rate of visual modality.

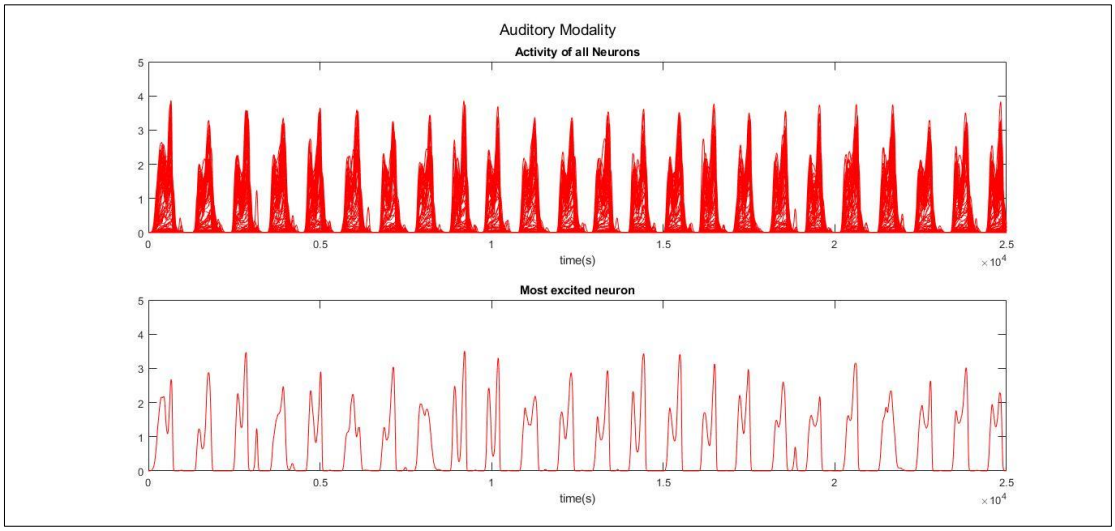

Fig. 5.25. Spike rate of auditory modality during attention modulation.

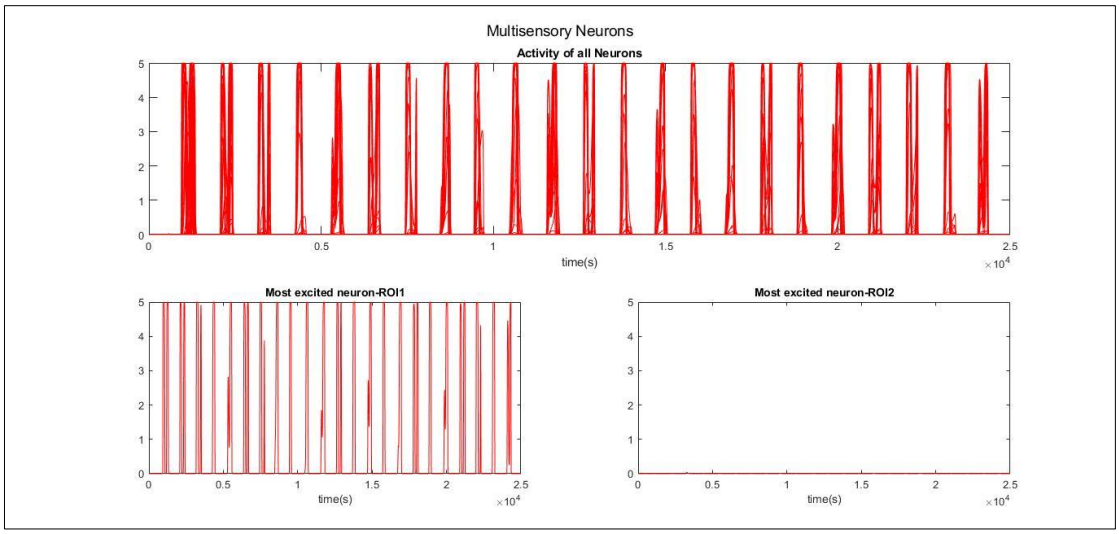

Fig. 5.26. Spike rate of multisensory area during attention modulation.

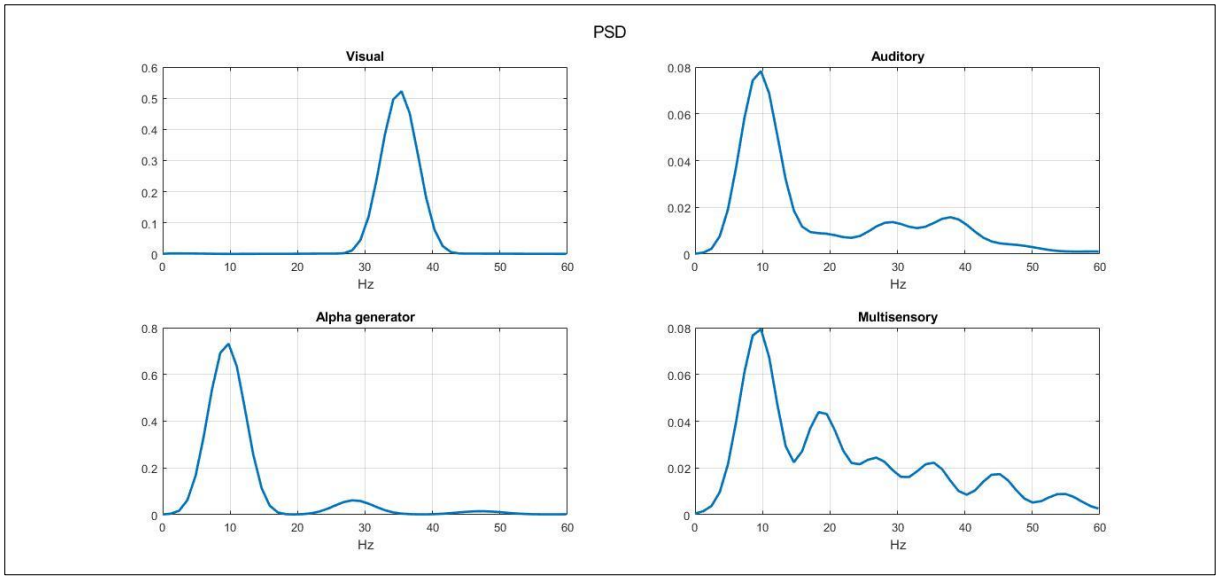

Fig. 5.27. PSDs of spike rates of all areas during attention modulation.

The oscillatory behavior of visual and auditory modalities can be seen in figures 5.24 and 5.25, respectively. The visual modality works in gamma band as usual while the auditory modality has gamma rhythm modulated by alpha. And for the multisensory area, figure 5.26 shows that indeed the neurons that are excited by the auditory input do not oscillate at all in this area. Further evidence can be found in figure 5.27 which depicts the PSDs of all four areas. From figure 5.25, in the auditory modality, the peak at 10Hz can be seen along with a very small peak at about 37Hz while no significant effect is observed in rest of the modalities' spectra.

The above results show how one modality can be entirely inhibited keeping the attention on the other. Another way to observe the mechanism of attention modulation is, for example, by inhibiting a certain spatial region in both modalities. In the following, we show inhibition of both unisensory modalities in the left half of azimuthal space i.e., 1^0-$90^{0.}$ The inputs for visual modality are applied at 60^0 and 120^0 while for auditory modality the inputs are provided at 40^0 and 140^0. Figures 5.28 and 5.29 show examples of positions (both unisensory and multisensory) without and with inhibition, respectively. Note that since inputs in both left and right side of space are quite close to each other they are identified as having one cause in the multisensory region, as depicted in figure 5.28.

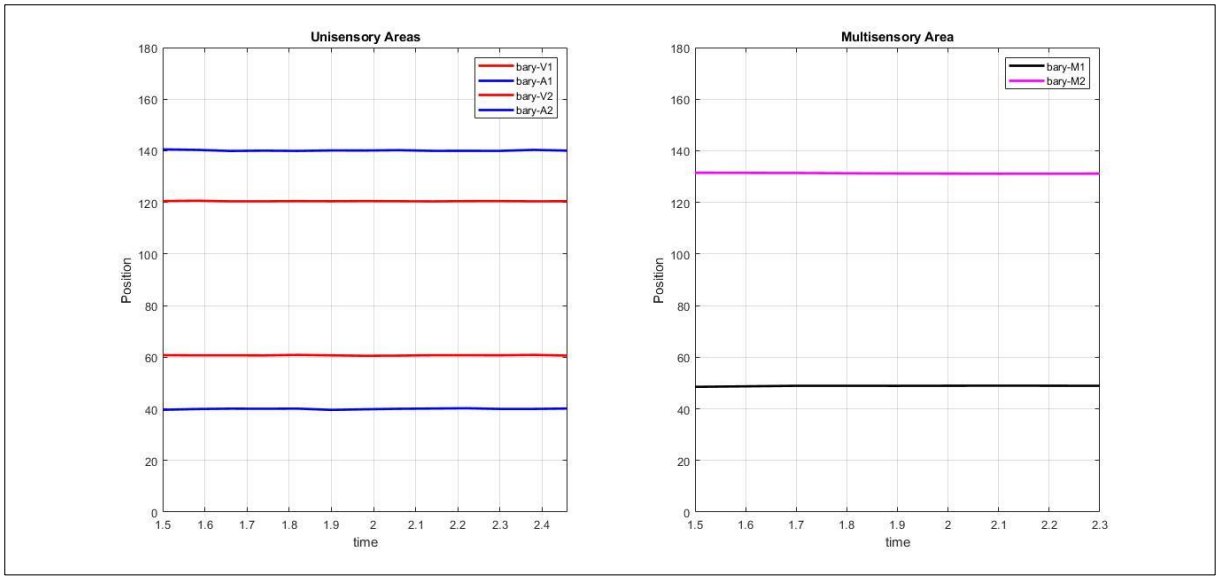

Fig. 5.28. Position information without inhibition.

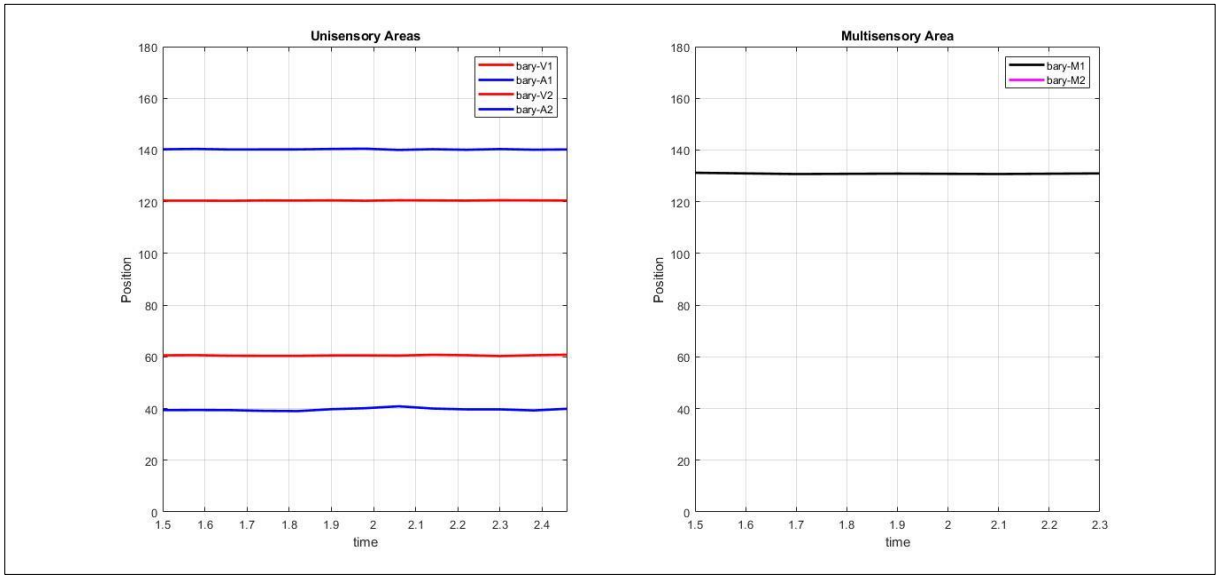

Fig. 5.29. Position information with inhibition.

The spike rate behavior of the above example is shown in figures 5.30 and 5.31 for both visual and auditory modalities, respectively. Since only the left half of azimuthal space is inhibited, the most excited neurons corresponding to left-side inputs are gamma-modulated-by-alpha while right-side inputs are only in gamma rhythm. The PSDs present the same information as in figure 5.27 so for brevity they are not shown here.

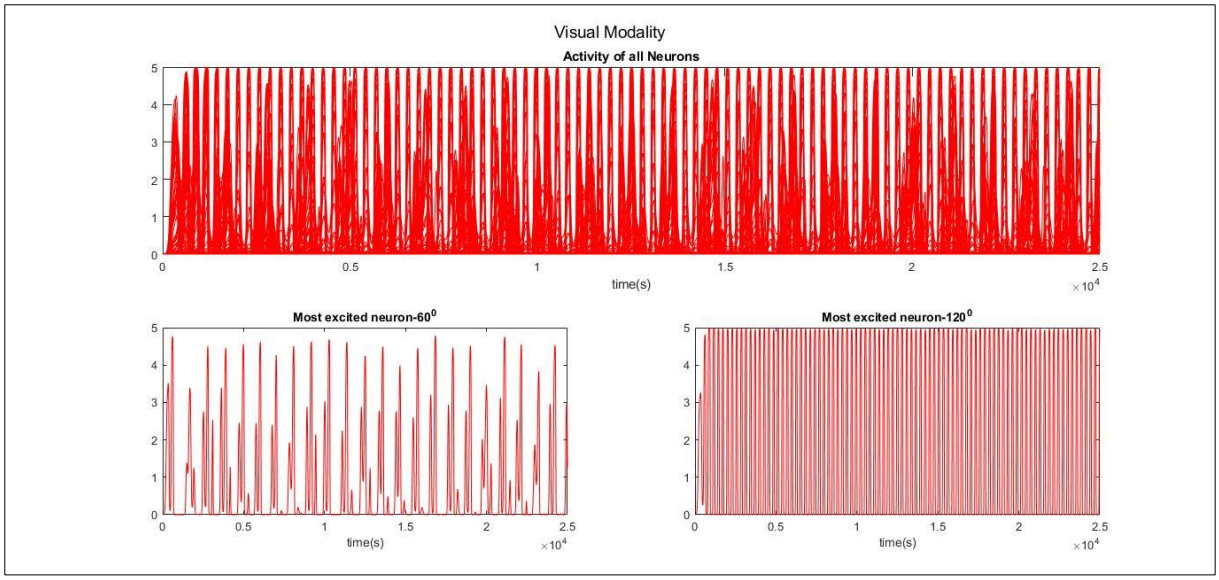

Fig. 5.30. Spike rate of visual modality during inhibition of left half of azimuthal space.

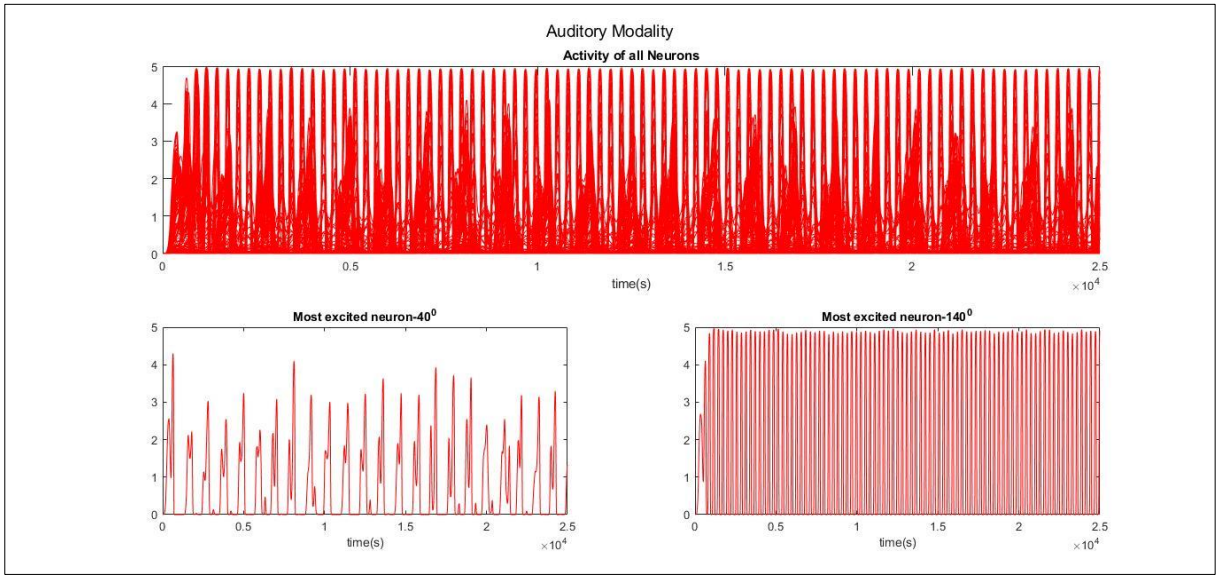

Fig. 5.31. Spike of auditory modality during inhibition of left half of azimuthal space.

Finally, we show the phase opposition of alpha rhythm between inhibited unisensory area compared to multisensory area; figure 5.32 depicts this example. We applied the visual and auditory inputs at 110^0 and 70^0, respectively, then inhibited the auditory modality. The activity of auditory modality is now, of course, in gamma modulated by alpha band which is the same as the multisensory area's activity but in opposite phase and therefore the activity is nullified. To compare the phases, we show, in figure 5.32, the activity of a single ROI (that was not inhibited) of multisensory area is shown along with the activity of highest excited ROI of auditory area. Both are in opposite phase and thus their combined effect is nulled out as shown in the last subplot of figure 5.26.

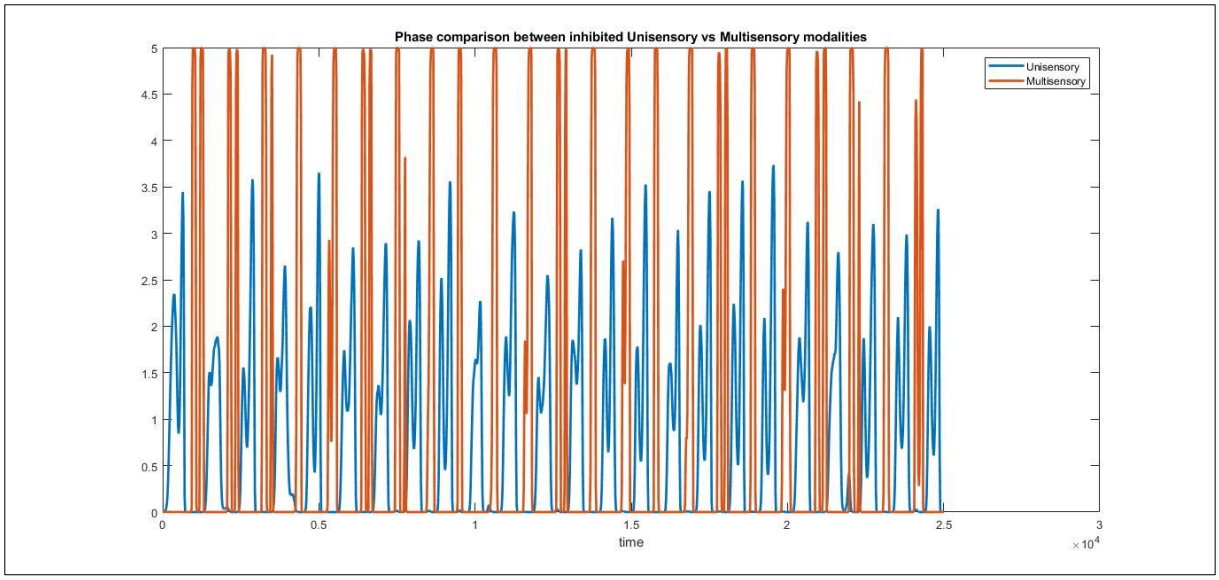

Fig. 5.32. Phase comparison between inhibited modality vs multisensory area.

Chapter 6 : Concluding Remarks

6.1. Discussion

There have been multiple approaches to model multisensory integration using different techniques such as Bayesian modeling (Ursino *et al.*, 2017), neural mass model (Moran and Reilly, 2006), and spiking neuron models (Lim, Keniston and Cios, 2011). The focus of this thesis is to use a neural mass model to realize audio-visual multisensory integration. Starting from a model of single ROI, to simulate oscillatory behavior of EEG in a certain frequency band, we moved forward to include 180 degrees of azimuthal space in each unisensory and multisensory modality. The model is then capable of simulating some of the physiological phenomenon pertaining to audio-visual multisensory integration such as ventriloquism, modulating attention, and causal inference.

The first aspect that deserves attention is the use of receptive fields (RFs) for respective unisensory modality. These receptive fields explain the excitation profile of all 180 ROIs given a spatial impulse at a certain position. After the ROIs are excited, they tend to oscillate in either alpha or gamma band or in gamma-modulated-by-alpha band depending on the choice of various parameters (as explained in chapter 4). The amplitude of spatial impulse was different for auditory modality compared to visual one. Now although most of the parameters for each modality are quite the same, since the RFs were different therefore a different value was needed to excite the ROIs of respective modalities. Furthermore, in the visual modality the amplitude of impulse was modulated by a linearly increasing function from the central position (90^{th} degree) towards either periphery.

This means that to excite the central region close to fovea a weaker input was required compared to peripheries where stronger input was needed to ensure the correct oscillatory pattern of ROIs. Physiologically it can be considered equivalent to, for example, more light being cast on an object that is present in the peripheral azimuthal space of eyes in order to excite ample number of retinal receptors for proper perception. While this is the case for visual modality, the auditory modality has quite uniform RFs throughout azimuthal space thus there is no need to vary the input amplitude.

Based on these inputs the excited ROIs started oscillating and from these oscillations we calculated the azimuthal position in both unisensory and multisensory areas. In the unisensory

areas the respective RFs are used to calculate positions, so we see that they are quite accurate. On the contrary, in the multisensory we simply used the average of both auditory and visual RFs to calculate position(s). We assumed that only the oscillatory information should be used to find position and not any a priori information about position from either unisensory areas. When calculating position in multisensory area, unisensory RFs in both modalities influence this calculation; since the RFs are different there is an obvious error in the multisensory position information. However, this error is not quite high. It follows a sinusoidal shape, as can be seen in figures 5.13 and 5.14, and is highest around 40 degrees away from central region for both modalities, standing at about 6-7 degrees.

Regarding the model's capability to infer causes/sources given multiple unisensory stimuli, the modal performed quite well in single modality compared to dual modality. In case of single modality, we provided two spatial impulses of the same unisensory modality at different positions. In the visual one, when one impulse was applied at the very center (90^0) and other impulse was varied, the modal could distinguish between causes when the azimuthal distance was about 27^0. This ability to distinguish causes decreased near periphery to 37^0 when the first impulse was shifted to 160^0 (see figures 5.17 and 5.18). This seems quite natural since the region near fovea consists of a dense population of neurons while sparsely populated neurons are found in peripheries.

Given the wider but uniform RFs of auditory modality, we only simulated one example of causal inference. One impulse was kept at 90^0 and the other one varied. To effectively distinguish between causes the auditory impulses needed to be at least 63^0 apart from each other (see figures 5.19 to 5.21). Finally, for the dual modality case, when one impulse in each unisensory modality was applied the model could distinguish causes when they were separated by 37^0 near central region and this distance decreased near peripheries to 24^0 (see figures 5.15 and 5.16). Also note that in case of unisensory conditions the unisensory position information was sufficient to infer causes while in dual modality case the multisensory position information was needed for causal inference.

These data were compared to the work on causal inference by (Cuppini *et al.*, 2017). The authors simulated a model for causal inference comprising two unisensory layers encoding auditory and visual stimuli and another downstream layer for multisensory integration which received connections from both unisensory layers. They showed that two unisensory inputs in visual modality could be distinguished (as having separate causes/sources) when they were

20^0 apart but did not talk about any differences occurring in central versus peripheral visual space. In the same paper, they showed that the same inputs but for auditory modality were perceived as having a single cause. Similarly, in the dual modality case, where one auditory and one visual input was applied at 20^0 distance from other, the model perceived it as having single cause. In case of visual modality our model performs almost the same while in auditory modality a larger distance is needed to distinguish between causes. This, of course, is due to wider auditory RFs.

The case of dual modality in our model is quite strange: In central region the ability to infer causes should be higher compared to peripheries but it is the opposite in our model. From figure 5.13 we see that in case of only visual input the barycenter of peripheral ROIS of multisensory area is shifted towards the center. Whereas in figure 5.14 we see the exactly opposite i.e., given only auditory input the barycenter of peripheral ROIs of multisensory area are shifted away from the center. Therefore, when peripheral inputs to both unisensory modalities are applied, these biases actually improve the causal inference ability compared to central region.

Regarding the ventriloquism effect, our model can only work in the spatial domain without considering the temporal aspects (like temporal ventriloquism such as sound-induced flash illusion). The model can simulate ventriloquism with maximum auditory shift occurring at 20^0 to 30^0 distance from the center (90^0) while for the visual modality the maximum shift is always less than 1^0. The data for visual modality is quite consistent with the data from previous work of (Ursino *et al.*, 2017), whereas the auditory modality performs slightly less. The authors showed that auditory modality shifted to a maximum of ~7^0 whereas in our model the auditory shift stands at about 3.8^0. One might suppose that this influence is proportional to the strengths of cross-modal connections so normally a higher strength used for cross-modal connection should lead to higher shifts. However, this was not the case for our model. Increasing the strength of cross-modal connections did provide a slightly higher shift but it was insignificant and further increase in strength only lead to loss of oscillations.

Finally, the model is capable of simulating attention: it can focus on a certain portion of azimuthal space whilst inhibiting the other, or it can inhibit one unisensory modality completely while focusing the attention on the other. This mechanism is realized in the multisensory area but not in the unisensory modalities. From figure 5.25 we see the activity of unisensory modality, in gamma-modulated-by-alpha rhythm, during inhibition. This activity is

now in opposite phase compared to the gamma-modulated-by-alpha activity of multisensory area as in figure 5.32. Therefore, when the multisensory area receives the inputs from both unisensory areas, the area whose activity is in opposite phase has its effect nulled out in the multisensory area as depicted in last subplot of figure 5.26. This is how attention mechanism is realized. In terms of human behavior this is equivalent to stating that the brain regions dedicated to unisensory processing process a number of stimuli all the time, but it is our conscious focus on a certain stimulus that leads to attention being modulated towards it while all other ambient information is inhibited. This focusing or attention process takes place in the multisensory areas.

6.2. Future Work

There are some aspects that need to be modified to achieve even better functioning of our model. These include,

1. Receptive Fields (RFs)
2. Cross-modal connections

The RFs were trained in a separate work by (Ursino *et al.*, 2017). The authors used a static model whereas our model is temporally dynamic (i.e., oscillatory) and thus in future we could train the RFs specifically tuned to our model which will ensure the proper excitation profile given a spatial impulse at certain position. Moreover, in the calculation of barycenter from oscillations we used preferred position of each neurons (denoted by K_j^s, where s=a,v and j ranges from 1 to 180). Of course, these preferred positions were calculated based on RFs and thus the unisensory position information was quite accurate. However, as we saw in the previous chapter, in the calculation of position in multisensory area there was some error. This error is attributed to improper use of both auditory and visual RFs when calculating position from oscillations. In future, we could train a separate RF for multisensory area which would have the influence of both unisensory RFs to ensure accurate calculation of position. It won't be easy since, of course, the RF of multisensory neurons varies depending on the presence of an auditory or visual input. Indeed, there are two RFs for each multisensory neuron. A way forward could be to calculate both visual and auditory RF for each multisensory neuron. Finally, the cross-modal connections too were trained using the static model by the same authors therefore, the next approach could be to train them based on our dynamic model.

References:

Barry E. Stein and Meredith, M. A. (1993) *The Merging of the Senses*. MIT Press.

Bartos, M., Vida, I. and Jonas, P. (2007) 'Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks', *Nature Reviews Neuroscience*, 8(1), pp. 45–56. doi: 10.1038/nrn2044.

Bastarrika-Iriarte, A. and Caballero-Gaudes, C. (2019) 'Closing eyes during auditory memory retrieval modulates alpha rhythm but does not alter tau rhythm', *NeuroImage*, 197(May 2018), pp. 60–68. doi: 10.1016/j.neuroimage.2019.04.053.

Burgess, A. P. and Ali, L. (2002) 'Functional connectivity of gamma EEG activity is modulated at low frequency during conscious recollection', *International Journal of Psychophysiology*, 46(2), pp. 91–100. doi: 10.1016/S0167-8760(02)00108-3.

Buzsáki, G. (2006) *Rhythms of the Brain*, *Rhythms of the Brain*. doi: 10.1093/acprof:oso/9780195301069.001.0001.

Cooper, N. R. *et al.* (2003) 'Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses', *International Journal of Psychophysiology*, 47(1), pp. 65–74. doi: 10.1016/S0167-8760(02)00107-1.

Cuppini, C. *et al.* (2017) 'A biologically inspired neurocomputational model for audiovisual integration and causal inference', *European Journal of Neuroscience*, 46(9), pp. 2481–2498. doi: https://doi.org/10.1111/ejn.13725.

David, O. *et al.* (2006) 'Dynamic causal modeling of evoked responses in EEG and MEG', *NeuroImage*, 30(4), pp. 1255–1272. doi: 10.1016/j.neuroimage.2005.10.045.

David, O. and Friston, K. J. (2003) 'A neural mass model for MEG/EEG: Coupling and neuronal dynamics', *NeuroImage*, 20(3), pp. 1743–1755. doi: 10.1016/j.neuroimage.2003.07.015.

David, O., Harrison, L. and Friston, K. J. (2005) 'Modelling event-related responses in the brain', *NeuroImage*, 25(3), pp. 756–770. doi: 10.1016/j.neuroimage.2004.12.030.

Diederich, A. and Colonius, H. (2004) 'Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time', *Perception and Psychophysics*,

66(8), pp. 1388–1404. doi: 10.3758/BF03195006.

Ergenoglu, T. *et al.* (2004) 'Alpha rhythm of the EEG modulates visual detection performance in humans', *Cognitive Brain Research*, 20(3), pp. 376–383. doi: 10.1016/j.cogbrainres.2004.03.009.

Ernst, M. O. and Bülthoff, H. H. (2004) 'Merging the senses into a robust percept', *Trends in Cognitive Sciences*, 8(4), pp. 162–169. doi: 10.1016/j.tics.2004.02.002.

Felleman, D. J. and Van Essen, D. C. (1991) 'Distributed hierarchical processing in the primate cerebral cortex', *Cerebral Cortex*, 1(1), pp. 1–47. doi: 10.1093/cercor/1.1.1.

Gazzaniga, M. S., Ivry, R. B. and Mangun, G. R. (2013) *Cognitive Neuroscience: The Biology of the Mind*. 4th edn. W. W. Norton & Company.

Gerstner, W. and Kistler, W. M. (2002) *Simplified spiking neuron models*.

Ghazanfar, A. A. and Schroeder, C. E. (2006) 'Is neocortex essentially multisensory?', *Trends in Cognitive Sciences*, 10(6), pp. 278–285. doi: 10.1016/j.tics.2006.04.008.

Gray, C. M. *et al.* (1989) 'Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties', *Nature*, 338(March), pp. 334–337.

Di Gregorio, F. *et al.* (2022) 'Tuning alpha rhythms to shape conscious visual perception', *Current Biology*, 32(5), pp. 988-998.e6. doi: 10.1016/j.cub.2022.01.003.

Grimbert, F. and Faugeras, O. (2006) 'Bifurcation analysis of Jansen's neural mass model', *Neural Computation*, 18(12), pp. 3052–3068. doi: 10.1162/neco.2006.18.12.3052.

Jansen, B. H. and Rit, V. G. (1995) 'Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns', *Biological Cybernetics*, 73(4), pp. 357–366. doi: 10.1007/BF00199471.

Jefferys, J. G. R., Traub, R. D. and Whittington, M. A. (1996) 'Neuronal networks for induced "40 Hz" rhythms', *Trends in Neurosciences*, 19(5), pp. 202–208. doi: 10.1016/S0166-2236(96)10023-0.

Jensen, O. *et al.* (2002) 'Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task', *Cerebral Cortex*, 12(8), pp. 877–882. doi: 10.1093/cercor/12.8.877.

Keil, A., Gruber, T. and Müller, M. M. (2001) 'Functional correlates of macroscopic high-

frequency brain activity in the human visual system', *Neuroscience and Biobehavioral Reviews*, 25(6), pp. 527–534. doi: 10.1016/S0149-7634(01)00031-8.

Kilpatrick, Z. P. (2013) 'Wilson-Cowan Model', *Encyclopedia of Computational Neuroscience*, pp. 1–5. doi: 10.1007/978-1-4614-7320-6.

Lim, H. K., Keniston, L. P. and Cios, K. J. (2011) 'Modeling of multisensory convergence with a network of spiking neurons: A reverse engineering approach', *IEEE Transactions on Biomedical Engineering*, 58(7), pp. 1940–1949. doi: 10.1109/TBME.2011.2125962.

Lopes da Silva, F. H. *et al.* (1974) 'Model of brain rhythmic activity - The alpha-rhythm of the thalamus', *Kybernetik*, 15(1), pp. 27–37. doi: 10.1007/BF00270757.

Ma, W. J. *et al.* (2009) 'Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space', *PLoS ONE*, 4(3). doi: 10.1371/journal.pone.0004638.

McGurk, H. and Macdonald, J. (1976) 'Hearing lips and seeing voices (McGurk Effect)', *Nature*, 264(5588), pp. 746–748.

Meredith, M. A. and Stein, B. E. (1983) 'Interactions among converging sensory inputs in the superior colliculus', *Science*, 221(4608), pp. 389–391. doi: 10.1126/science.6867718.

Meredith, M. A. and Stein, B. E. (1986) 'Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration', *Journal of Neurophysiology*, 56(3), pp. 640–662. doi: 10.1152/jn.1986.56.3.640.

Moran, R. J. and Reilly, R. B. (2006) 'Neural mass model of human multisensory integration', *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pp. 5559–5562. doi: 10.1109/IEMBS.2006.259588.

Rodriguez, E. *et al.* (1999) 'Perception's shadow: Long-distance synchronization of human brain activity', *Nature*, 397(6718), pp. 430–433. doi: 10.1038/17120.

Rotterdam, A. van *et al.* (1982) 'A model of the spatial-temporal characteristics of the alpha rhythm', 44(2), pp. 283–305. doi: https://doi.org/10.1016/S0092-8240(82)80070-0.

Rowland, B. A. *et al.* (2007) 'Multisensory integration shortens physiological response latencies', *Journal of Neuroscience*, 27(22), pp. 5879–5884. doi: 10.1523/JNEUROSCI.4986-06.2007.

Shams, L; Kamitani, Y.; Shimojo, S. (2000) ‘What you see is what you hear’, *Nature*, 408(December), p. 788.

Sotero, R. C. *et al.* (2007) ‘Communicated by Olivier Faugeras Realistically Coupled Neural Mass Models Can Generate EEG Rhythms’, *Neural Computation*, 512, pp. 478–512. Available at: https://www.mitpressjournals.org/doi/pdf/10.1162/neco.2007.19.2.478.

Stanford, T. R. and Stein, B. E. (2007) ‘Superadditivity in multisensory integration: Putting the computation in context’, *NeuroReport*, 18(8), pp. 787–792. doi: 10.1097/WNR.0b013e3280c1e315.

Stein, B. E. and Stanford, T. R. (2008) ‘Multisensory integration: Current issues from the perspective of the single neuron’, *Nature Reviews Neuroscience*, 9(4), pp. 255–266. doi: 10.1038/nrn2331.

Tononi, G. and Edelman, G. M. (1998) ‘Consciousness and complexity’, *Science*, 282(5395), pp. 1846–1851. doi: 10.1126/science.282.5395.1846.

Ursino, M. *et al.* (2007) ‘Use of a neural mass model for the analysis of effective connectivity among cortical regions based on high resolution EEG recordings’, *Biological Cybernetics*, 96(3), pp. 351–365. doi: 10.1007/s00422-006-0122-4.

Ursino, M. *et al.* (2017) ‘Development of a bayesian estimator for audio-visual integration: A neurocomputational study’, *Frontiers in Computational Neuroscience*, 11(October). doi: 10.3389/fncom.2017.00089.

Ursino, M., Cona, F. and Zavaglia, M. (2010) ‘The generation of rhythms within a cortical region: Analysis of a neural mass model’, *NeuroImage*, 52(3), pp. 1080–1094. doi: 10.1016/j.neuroimage.2009.12.084.

Ward, L. M. (2003) ‘Synchronous neural oscillations and cognitive processes’, *Trends in Cognitive Sciences*, 7(12), pp. 553–559. doi: https://doi.org/10.1016/j.tics.2003.10.012.

Wendling, F. *et al.* (2000) ‘Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals’, *Biological Cybernetics*, 83(4), pp. 367–378. doi: 10.1007/s004220000160.

Wendling, F. *et al.* (2002) ‘Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition’, *European Journal of Neuroscience*, 15(9), pp. 1499–1508. doi: 10.1046/j.1460-9568.2002.01985.x.

White, J. A. *et al.* (2000) ‘Networks of interneurons with fast and slow γ-aminobutyric acid type A (GABA(A)) kinetics provide substrate for mixed gamma-theta rhythm’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(14), pp. 8128–8133. doi: 10.1073/pnas.100124097.

Wilson, H. R. (1999) *Spikes, Decisions, and Actions: The Dynamical Foundations of Neuroscience*. Oxford University Press.

Wilson, H. R. and Cowan, J. D. (1972) ‘Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons’, *Biophysical Journal*, 12(1), pp. 1–24. doi: 10.1016/S0006-3495(72)86068-5.

Zavaglia, M. *et al.* (2008) ‘The effect of connectivity on EEG rhythms, power spectral density and coherence among coupled neural populations: Analysis with a neural mass model’, *IEEE Transactions on Biomedical Engineering*, 55(1), pp. 69–77. doi: 10.1109/TBME.2007.897814.