

Extension of Superior Colliculus Based Models
of Multisensory Integration to the Cortex
Michael Hadley
Lafayette College

Abstract:

The field of multisensory integration lacks a comprehensive theory of how information from the different senses can be integrated. Multiple models have been proposed, but each only shines light on a piece of the puzzle. There is no common ground upon which all the models can build. A SOM based model of multisensory integration is a double-edged sword. On the one hand, the model is quite transparent. It is easily predictable what will happen given initial conditions. On the other hand, the MSE is dependent on the parameters- in particular the sigmoid transfer function.

I was able to extend the model to more complex senses with the same results, but uncovered some deficiencies. The settings for parameters become much more unclear as you move up in complexity. SOMs do not “magically” integrate information, but rather work on very simple principles to categorize data along a particular dimension. A single two dimension map cannot handle more than one relationship among the data (i.e. it cannot organize data alphabetically and numerically). I propose a testable model to move the SOM based models to the next level.

Extension of Superior Colliculus Based Models of Multisensory Integration to the Cortex

1. Introduction

In reading this proposal, your eyes and visual system are deciphering the words on this page, and the touch receptors on your skin inform you that you are indeed holding a paper in your hands (or that you are scrolling through this screen). Simply eating an apple involves vision, taste, touch, hearing and smell. The world is intrinsically multisensory, and our brain handles it as such. Our sensory experiences are so intertwined that we cannot consciously fully separate them out; doesn't that suggest that our brains process the information from all the senses as a whole rather than parts?

Imagine that shiny, red apple again. I hold it out to you, and I let you take a bite. Experience the sweet, tart taste. I want you to only experience the apple as an entity that only has taste properties. Now, I start tossing the apple up in the air and catching it. Think you still are only experiencing taste? Suddenly, I throw the apple at your head. What do you do? Duck, move, shield your face, catch the apple, etc. Your brain never forgot that the apple has physical properties other than taste.

Scientists have traditionally studied the senses in isolation. As a result, the standard view of sensory processing in the brain is that the senses process their own information separately, and then at a much higher level of processing, the sensory information is combined. It has only been very recently that this view been challenged and the senses viewed as an integrated whole. New studies show both anatomical and functional connections between low-level sensory areas in the cortex (Ghazanfar & Schroeder, 2006).

While some work has been done computationally modeling multisensory integration in subcortical¹ areas responsible for gaze orientation, multisensory integration in the cortex has remained untouched. Despite the sudden surge in studies of multisensory integration in the cortex, no cohesive theory of cortical integration has been accepted. Arguable the biggest problem facing the subfield of cortical integration is this lack of a theory with testable predictions. The cortex is an expansive area with an estimated 20 billion neurons². New technology will never diminish the number of neurons we would have to sort through to find the regions and pathways involved in integration. Without predictions and direction, our technology is not being efficiently employed. Computational modeling can bring together the lines of evidence and make new, testable predictions. Computational modeling can direct our investigation into the human neural architecture.

My proposal is to apply the important facets of the subcortical computation to a cortical setting. As reviewed below, the cortical integration as currently understood has features similar to the process in subcortical systems and given evolution's preference for reusing mechanisms (multisensory integration occurs in organisms as closely related as monkeys and as distant as the fruit fly), it seems likely that similar mechanisms might be operating in both subcortical and cortical regions. While integration has been demonstrated in the much simpler subcortical regions using neural recording techniques, the data is sparser for cortical regions. Therefore my model will be mostly theoretical in nature. It will be aimed at demonstrating integration as

¹ A major division in the brain is subcortex/cortex. The "lower" subcortical regions evolved first and are thought to perform more basic functions like breathing and coordination of movement. The "higher" cortical regions evolved later and are through to be responsible for our more sophisticated behaviors like planning and memory. (<http://faculty.ed.uiuc.edu/g-cziko/wm/05.html>)

² <http://vadim.oversigma.com/MAS862/Project.html>

defined by the subcortical literature while still remaining congruent with the existing cortical evidence.

Understanding multisensory integration in the cortex would have direct and important implications for learning, sensory loss and sensory substitution. Education has recognized the importance of multisensory cues in learning for years (Shams & Seitz, 2008), but with the mechanisms behind the integration of those cues understood, we may be able to develop pedagogy to learn better and faster. It's already been shown that multisensory cues allow six month old infants to perform numerical computations that were previously thought impossible for their age (Jordan, Suanda & Brannon, 2007). Many people wish to learn a language, but just do not have time or energy to actually carry through. Understanding multisensory integration could make the process much more digestible for our system. Hearing and seeing words at the same time may require less conscious effort and time on our part to memorize vocabulary, grammar rules, etc.

Sensory loss occurs when either a sensory organ (like the eyes, tongue, ears, etc.) or the sensory machinery in the brain is damaged or does not develop properly. We have very limited capabilities to help people compensate for sensory loss because we lack an understanding of how the sensory organs interface with the brain and other sensory organs. One field that has attempted to restore lost senses is sensory substitution. The general idea is to hijack a sense to process information from another modality. One example most people are familiar with is Braille: ordered bumps that represent letters and words. Another more recent development is auditory-visual conversions. Through the application of an algorithm called vOICE, visual information as captured by a camera can be converted into auditory information. Individuals who have lost their sight can "see" using such an apparatus (Kim & Zatorre, 2008; Bach-y-Rita & Kercel, 2003). New insights into multisensory integration could lead to a future where we simulate new sensory structures; for example, a nightvision or thermal camera could just as easily be used within the sensory substitution technology.

2. The Senses

We have at our disposal seven senses: vision, smell, taste, touch, hearing, balance and kinesthesia³. Each sense lets us understand a tiny slice of the world around us. Each of the basic senses uses specialized structures (sensory receptors) in their sensory field to collect physical information from the world and converts it into neural firing patterns. The sensory receptors for the senses are distributed all over the body, but the information from the receptors is all sent to the brain. Figure 1 shows various brain structures associated with the three main senses- vision, hearing and touch. Primary areas receive raw data from the sensory receptors and transform/analyze that data. The output of primary areas is sent to secondary areas for more transformation/analysis.

The brain converts informational forms as disparate as light and pressure to a common: format- action potential firing rates. Once information enters the brain, any neurological computation can be applied. My proposed model follows the subcortical model's assumption that the senses only differ in content not in form. This is supported by literature on plasticity and topography.

The similarities in processing in sensory areas are apparent. Starting with development, all of the senses areas are determined by a combination of genetics and environment. Almost

³ <http://academics.tjhsst.edu/psych/oldPsych/senses/>

every person has two eyes, a visual brain area, two ears, an auditory brain area, etc. Still, not everyone has the same perceptual experiences of the world. Visual perception across cultures has been well studied. School-aged children of Ghana that grew up in environments relatively devoid of pictures were quite hard pressed to recognize depth in pictures. In fact, almost none of the children studied succeeded in using the depth cues (Robert, 1973). Take a group of similar aged group western children. You would be hard pressed to find a child that did not use the depth cues. Both genes and experience influence the development of the senses.

The key process that allows experience to influence the structure of sensory areas is neural plasticity⁴. Plasticity seems to operate in the same way in all cortical regions. Rauschecker (1999) suggests that findings on the topic of auditory cortical plasticity can really be applied to any sensory system. He suggests that the same plasticity rules that allow us to learn the phonetics of language operate in the other senses.

Not only are the rules of learning (plasticity) the same across senses, but the structure of what is learned is the same. There are detectable, topographic maps for each sensory system (retinotopy for the eyes, somatotopy for the skin, etc) (Killackey, Rhoades & Bennett-Clarke, 1995). Psychophysical, neurological and neurophysiologic mechanisms underlying detection and representation in the senses are very similar (Wagner, Kautz & Poganiatz, 1997). Kahonen and Hari found (1999) the same ordering underlying the senses and suggest that order is developed by a common algorithm (self-organizing map).

This evidence is preliminary, but it suggests that the only differences between the senses are the sensory organs (eyes versus ears) and the content (visual information like brightness versus auditory information like loudness). The senses all seem to operate under the same learning rules. Additionally, they all appear to organize themselves in topographic maps to represent content. The simplification that the only difference between the senses is content is currently justified.

Another simplification of both my model and the subcortical models is a focus on integration between visual, auditory and somatosensory sensory information. Although there is some evidence to suggest that the underlying mechanisms of the other senses are similar (Small, 2004), the bulk of the literature speaks directly to visual, auditory and somatosensory systems.

3. Multisensory Integration

The literature on multisensory integration spans multiple disciplines and techniques. First, there are behavioral studies that show the benefits of having multisensory cues available in a memory or reaction time task. Second, there are single-cell neural firing rate recordings from subcortical areas showing multisensory integration. Finally, there has been a recent explosion of work in the cortical regions using single-cell recordings, neural tracing and neural population recording techniques.

3.1. Psychology

What is a sufficient definition of multisensory integration? It depends on context, on how the term is being applied. A very general definition would be the process by which sensory information from multiple modalities is used collectively to create a more veridical representation of the world. In other words, the sensory information from one sense is combined with the information from another sense in a way that reduces variance as compared to each

⁴ The ability of a network to change its architecture in response to new experiences or brain damage

sense's variance (Ernst & Bühlhoff, 2004). This definition of multisensory integration works well in the context of psychological research, which is the first area of evidence for multisensory integration in our brains.

We live in a multisensory world; it is obvious to us on personal level that we construct our world from all our senses. It seems logical to suppose that the brain must use the information from our senses together not separately. Apart from personal experience, psychological experiments are the best way to confirm the importance of multisensory integration in our brains. By figuring out how multisensory integration works at a high level, we can postulate how neurons might accomplish the task. The general principle is that if something is significant at the high level, there are most likely significant things at a lower level (Driver & Noesselt, 2008). Psychological research can direct more invasive research.

The first piece of evidence comes from multisensory illusions. A popular one is called the rubber hand illusion⁵. As in Figure 2, a rubber hand (which is in view of a person) is placed next to their real hand (which is hidden from view). Both the real hand and the rubber hand are stimulated with a brush in the same place at the same time. If this is continued for long enough, most people will identify the rubber hand as somehow part of their self. Interestingly enough, if you ask them to close their eyes and point at their hand, they point at the rubber hand. The synchronous visual and tactile information causes the brain to accept the rubber hand as part of self. It's a fairly logical mechanism. If you are feeling and seeing a stimulus on a hand-like object in a region where you believe your hand is currently located, almost every single time this means that hand-like object is your hand. These types of illusions show that information enters the brain from multiple pathways, and the brain creates an estimate of the world based on all available sensory information (Driver & Noesselt, 2008).

Numerous behavioral paradigms have shown the benefits of having multisensory information available (Olivers & Van der Burg, 2008; Shams & Seitz, 2008; Collignon et al., 2008; Jordan, Suanda & Brannon, 2007; Suied, Bonneel & Viaud-Delmon, 2009; Philippini, van Erp & Wekhoven, 2008). Shams and Seitz (2008) give a typical experiment. Two groups of human participants were asked to perform a task that required some learning. Both groups were trained using visual stimuli. The difference between the groups was that the second group also received auditory stimuli during training whereas the first group received none. The task performed only directly required learning in the visual modality (e.g. recalling whether a picture was in the training set or not). Audio-visual participants learned as much as 60% faster and performed better than their solely visual counterparts. Collignon et al. (2008) confirmed these results and tested how attention affects integration. Even if participants were told to only pay attention to information in one modality, the "ignored" information influenced performance.

Olivers and Van der Burg (2008) replicated a similar experiment with some more controls. When two visual targets are presented in rapid succession, the second of the two is very often missed (and this is called attentional blink). Presenting a synchronous sound prevented attentional blink. If the sounds were presented just before the second visual stimuli, attentional blink persisted. The study suggests the integration of multisensory cues is automatic and not due to alerting (Olivers & Van der Burg, 2008). The effects of multisensory cues persist even with more complex tasks (Philippini, van Erp & Wekhoven, 2008). Also, the effects of multisensory cues can be seen in children as young as six months old (Jordan, Suanda & Brannon, 2007).

⁵ <http://www.newscientist.com/article/dn16809-body-illusions-rubber-hand-illusion.html>

A few recent studies show the importance of congruence of multisensory cues in enhancement of learning (Collignon et al., 2008; Shams & Seitz, 2008; Suied, Bonneel & Viaud-Delmon, 2009). Congruency is defined as the relationship between sets of sensory cues in prior experience. Seeing a phone and hearing a ring is congruent since this is very common in our experience. Seeing a phone and hearing the sound of running water is incongruent because experience (hopefully) does not connect these two cues. The general finding is that congruence produces better performance on multisensory tasks than their associated unisensory tasks and incongruence shows lower performance on multisensory tasks (Collignon et al., 2008; Shams & Seitz, 2008). The results of a study by Suied, Bonneel & Viaud-Delmon (2009) suggest that the senses involved in the congruency relationship change the results. There is also evidence that suggest an artificial relationship that is initially incongruent can become congruent through learning and then enhance performance (Shams & Seitz, 2008).

The field of psychology has shown that availability multisensory information speeds up our processing of the world and/or makes our decisions more accurate depending on the situation. These processing features found in humans are probably evolutionarily conserved since similar results were found with *Drosophila* with a visual and olfactory task setup (Shams & Seitz, 2008). A more thorough analysis from the psychological perspective is still needed. Our knowledge of how context, timing, spatial organization of multimodal stimuli effects multimodal integration is lacking (Driver & Noesselt, 2008). Congruence seems to be a very important factor, and the evidence suggesting congruence can be learned may mean that the mechanism is just about associations, in other words, the more associations in your brain for a particular idea, the faster/better the recall of that idea.

3.2 Subcortical Areas

Guided by the assurances that multisensory integration exists and has a real effect on behavior, we shift our focus to the neuronal level. First, we need a new definition of multisensory integration that is consistent with the psychological findings but that better captures the process at the neuronal level. The accepted definition is: a statistical significant difference between the mean firing rate of a neuron in response to a multimodal stimuli and the mean firing rate of the most effective component unimodal stimulus. If the multimodal stimulus is auditory (A) and visual (V), the mean firing rate of a neuron in response to the AV is compared to the greatest mean firing rate when presenting A or V individually.

A summary of the different types of multisensory integration can be seen in Figure 3. If the multisensory response is greater than the largest unisensory response, this is called multisensory enhancement. If the multisensory response is less than the largest unisensory response, this is called multisensory depression. If there is no significant difference between the multisensory response and the largest unisensory response, there is no multisensory integration (Stein et al., 2009).

Within multisensory enhancement, there are three subclassifications. If the multisensory response is greater than the sum of the unisensory responses, it is considered superadditive. If the multisensory response is less than the sum of the unisensory responses, this is called subadditive. If there is no difference between the sum of the unisensory responses and the multisensory response, it is called additive (Stein et al., 2009).

Within multisensory depression, there are also three subclassifications. If the multisensory response is greater than smallest unisensory response, it is considered superminimal. If the multisensory response is less than the smallest unisensory response, this is

called subminimal. If there is no difference between the sum of the smallest unisensory response and the multisensory response, it is called minimal (Stein et al., 2009).

The most thoroughly studied brain area in the multisensory integration field is the superior colliculus (SC). The SC is responsible for detecting sensory targets and orienting towards them. It is a subcortical midbrain structure that mainly receives unisensory afferents of two kinds: ascending and descending. The ascending connections come from the visual, auditory and somatosensory systems. The descending connections come from cortical regions like the parietal cortex in primates (Anastasio & Patton, 2003). In cats, the descending connections come from other cortical regions: the anterior Ectosylvian sulcus and the rostral portion of the lateral suprasylvian sulcus (Rowland, Stanford & Stein 2007).

Early single-cell recordings⁶ of the deep SC layers (SC) showed that many of the neurons responded to inputs from multiple senses (Stein et al., 2009). In cats, about one-half of the SC neurons respond to multiple senses while in monkeys, about one-fourth respond to multiple senses (Anastasio & Patton, 2003). The vast majority of these “multisensory” neurons can actually integrate multisensory information⁷ (Stein et al., 2009).

It has recently been found that when the descending cortical afferents are ablated or temporarily deactivated, the SC neurons no longer integrate multisensory information yet they still are responsive to multiple senses. In other words, a particular neuron that responded to audio, visual and somatosensory inputs no longer has a mean multimodal firing rate that is significantly different from the greatest mean unisensory firing rate (Rowland, Stanford & Stein, 2007). Ablation/disabling of the cortical afferents causes animals to be unable to integrate multisensory information for orientation and localization behaviors, but they retain their ability to respond to any particular modality-specific stimulus. Similar results are found when an animal is reared in the dark i.e. deprived of visual input and any multimodal inputs involving vision.

Based on the ablation evidence, the SC could not be implementing a linear computation to integrate sensory information. A linear model would predict that removing certain inputs should change unisensory and multisensory responses in same way. Instead, the response to unisensory is intact while the response to multisensory is destroyed (Stein et al., 2009). This has lead multiple authors to speculate that the cortical afferents must play a critical modulatory role (Rowland, Stanford & Stein, 2007; Anastasio & Patton, 2003)

Single neuron recordings in the SC have also uncovered a relationship between unisensory and multisensory responses called inverse effectiveness. Experiments showed that the smaller the unisensory response, the greater the magnitude of multisensory response (and the larger the unisensory response, the smaller the magnitude of multisensory response). For multisensory enhancement, this means the smaller the unisensory response, the greater the positive multisensory response. For multisensory depression, this means the smaller the unisensory response, the greater the negative multisensory response. Figure 4 shows a typical study's results. The first graph shows the response to varying intensities of a visual stimulus. (The auditory stimulus remains constant.) The second graph shows the multisensory response corresponding to each intensity of visual stimulus. The final graph shows the inverse

⁶ A single-cell recording is performed by inserting an electrode into a neuron. This electrode measures electrical activity which can be translated into firing rate information.

⁷ Recall that while responding to multiple senses is necessary for multisensory integration, it is not sufficient. A neuron can respond to multiple senses, but not show any statistical difference when comparing the mean multisensory firing rate to the greatest mean unisensory firing rate.

effectiveness trend: AV1 shows a 180% increase in response as compared to V1, but AV5 only shows a 58% increase in response as compared to V5.

One interpretation of inverse effectiveness is in terms of information. The stronger the unisensory responses, the more information contained in each unisensory response, and the less information is gained from combining the unisensory information. Weaker unisensory responses indicates less information available from each sense, and hence, more information is gained from combining the unisensory information. This implies that when the unisensory responses are strong, their information becomes redundant to a certain extent. The weaker the unisensory responses, the less redundant the information is (Stein et al., 2009).

It is important to note that while inverse effectiveness seems to be a general rule both at the level of single neurons and the level of populations of neurons, there are exceptions. Some studies show that increasing the intensity of unisensory responses did not decrease multisensory enhancement, but actually increased it. Stein et al. (2009) suggest this may be because the cross-modal stimuli are bound more effectively with increasing unisensory response. Also, there have not been studies explore the correlates of inverse effectiveness at the behavioral level.

The last important observation comes from studying the development of multisensory integration in the SC. Wallace and Stein (1997) performed single and multi-unit recordings from neurons in the SC of kittens ages 3 days to 135 days. No multisensory neurons were found until 10 days postnatal. These early multisensory neurons had little or no integrative ability (i.e. no significant response enhancement or depression). Suddenly around day 28 postnatal, the multisensory neurons began integrating sensory cues. Some of the properties of these early integrative neurons are indistinguishable from the adult integrative neurons: inverse effectiveness, superadditive response, and magnitude of response enhancements. The key difference between early and adult integrative neurons was adult neurons have a temporal window over which they can integrate sensory cues (about 250 msec), but early neurons required essentially simultaneous presentation of sensory cues. Still, the SC is not considered to be adult with respect to integrative ability until about three months postnatal (Wallace & Stein, 1997).

The kitten results parallel some results from studies on children. It is known that adults show statistically optimal multisensory integration, but Gori et al. (2008) found that children under the age of eight show integration that is very far from statistically optimal. Around the age of eight, suddenly children show statistically optimal integration revealing a shift from relying on the senses independently to integrating them together. They tested children of ages in the range five through ten on two tasks: size discrimination and orientation discrimination. Gori et al. (2008) propose that the lack of optimality in integration is because the bodies of eight to ten year olds are still growing significantly. Eyeball length, length of digits, size of limbs, etc. are all changing. The nervous system must constantly recalibrate the senses individually before they can be integrated (Gori et al., 2008).

This developmental evidence should be taken as preliminary. This is the first study to investigate how integrative ability develops in humans, and it only tested children in two specific tasks. Multisensory integration is a complex phenomenon that is dependent on context, as the psychological evidence suggests.

The SC studies provide a wealth of data that can inform computational models of multisensory integration. Single cell recordings give a basis for analysis the firing patterns of artificial neurons. The ablation studies suggest the role of descending connections is modulatory. Developmental studies suggest that there is a critical developmental point in which a structure or functional change suddenly turns multisensory responsive neurons into

multisensory integrative neurons. Inverse effectiveness shows some of the effects of context on integration, but this is an area that still needs much work. The role of context/relationship in multisensory integration is still poorly understood.

3.3 Cerebral Cortex

Cortical multisensory integration is finally becoming a focus of research attention. The standard theory that the senses are independent until higher areas of the cortex was motivated by early studies showing little or no connection between the low-level sensory systems and lesion studies that apparently showed unimodal behavioral defects. Recently, studies have challenged these initial assumptions (Ghazanfar & Schroeder, 2006). There are two lines of evidence: structural and functional.

The structural evidence comes from tracing studies. Falchier et al. (2002) performed a tracing study of the monkey primary visual cortex (V1) using sensitive tracers aimed at the areas of V1 responsible for peripheral vision. They found that the core and parabelt regions of the auditory cortex and the association area STP all have afferents in V1 (Falchier et al., 2002). Rockland and Ojima (2003) also reported connections from auditory association areas to areas of V1 and V2 responsible for peripheral vision. They also suggest that these connections may have been overlooked because the connections are sparse, and older tracing techniques may not have been able to distinguish the connections from background noise. Studies on individuals who are born deaf show enhanced peripheral visual attention compared to their central visual attention (Bavelier, Dye & Hauser, 2006). This may suggest an important inhibitory role of the connections between auditory areas and V1.

The idea that auditory areas and peripheral vision are connected has interesting implications. It suggests that auditory areas may be supplementing the peripheral vision to create a more accurate picture of the world. There may be more directed connections between low-level regions of other senses. For example, vision may make up for areas of the skin that lack spatial resolution. Other studies have found somatosensory inputs to the caudomedial belt region including other belt and parabelt regions (but not A1). Current models of the organization of our sensory systems suggest that signal can be carried to almost any cortical area and most subcortical areas (Schroeder et al., 2003).

The evidence from anatomy is incomplete because the existence of connections does not suggest how those connections might be used. We gain little insight into the purpose of these connections (if any) just by their existence. The studies of function supplement the anatomical findings. Figure 5 presents a summary of the functional evidence for crossmodal integration at multiple levels of brain processing. Figure 5a shows the traditional regions believed to be multisensory, while Figure 5b represents the areas that are now thought to be multisensory (Ghazanfar & Schroeder, 2006). The data that is represented in Figure 5b comes from three main sources: single-cell recordings showing integration; recording techniques showing multisensory convergence⁸; population recordings showing integration.

Few studies have performed single single-cell recordings of higher brain regions. A recent study found that 23% percent of the neurons in the superior temporal sulcus (a higher cortical region) showed multisensory integration. The study also found that this integration was dependent on congruence of the stimuli from multiple modalities (Ghazanfar & Schroeder,

⁸ Multisensory convergence is defined as a brain region being responsive to multiple modalities. This is necessary, but not sufficient for multisensory integration. It is generally assumed in the field that convergence most likely indicates integration (Stein & Stanford, 2008).

2006). The cat AES (a cortical region) has neurons that do not project to the SC that integrate information in the same way that SC neurons integrate information. The spatial and temporal dynamics match as well as the principle of inverse effectiveness applies to both (Stein & Stanford, 2008).

Many studies simply show a convergence of multisensory information (Stein & Stanford, 2008). A study using a microelectrode mapping (multiple single-cell recordings) of lower sensory areas examined the somatosensory inputs to the auditory areas. They appear to be cutaneous representations of the head and neck area. This fits with the proposed function of those auditory areas as holding additional spatial information (Schroeder et al., 2003). One study used cellular recordings to show that the posterior parietal cortex (PPC) receives signals from many modalities and combines them through a gain field mechanism. This study does not adhere to the standard definition of multisensory integration, so all that can be said is that the area is multisensory. The PPC is believed to be responsible for holding a multimodal mapping of space (Andersen et al., 1997). Areas of the frontal cortex and premotor cortex have neurons that respond to multiple senses, but again, this is not the same as integrating the senses. Single cell recordings in the auditory cortex of the macaque monkey confirm multisensory convergence (Ghazanfar & Schroeder, 2006).

Population recording techniques look at a cluster of nearby neurons. While the technique has the advantage of looking at larger patterns in the brain than single-cell recording, there are some severe limitations. When using a population technique, the activity of large numbers of neurons is collapsed blurring the line between integration and regular activity. The responses of a population that integrates information may be washed out by surrounding populations that do not integrate. Conversely, a population study may collapse multiple unimodal populations into a single “multimodal” population. Given these limitations, it is surprising that there is no standard definition of integration across these studies. A great many use the condition that the population response must be superadditive (i.e. the population’s response to the multisensory must be greater than the sum of the population’s response to the unisensory components). Any neurons that fit this definition are integrative, but it excludes many forms of integration. Recent cellular work suggests that linear summation is a common mechanism of multisensory integration; these cells would not be multisensory according to this definition. Others have used the max or mean criteria. The max criterion means that the multisensory response just needs to be greater than the greatest of the unisensory responses. The mean criterion says that the multisensory response must be greater than the mean of the unisensory responses. Each definition has its advantages and disadvantages. Some consensus needs to be reached so the results from multiple studies can be compared (Driver and Noesselt, 2008).

Ghazanfar et al. (2005) studied multisensory integration in the auditory cortex using intracranial electrodes to record local field potentials. Specifically, they inserted electrodes into the core (primary and primary-like auditory processing) and lateral belt region (higher-order auditory processing). These electrodes are placed along the surface of these regions in order to record the overall response without singling out any particular neural response. Their study is special because they used more realistic stimuli than typically used; they displayed movies of a macaque grunting or cooing paired with audio of a macaque grunting or cooing. They found both multisensory enhancement and depression in both regions although it was stronger in the lateral belt region. The study used the standard single neuron definition of multisensory integration (a statistical difference from the greatest unisensory response). They also found that the responses matched the principle of inverse effectiveness (Ghazanfar et al., 2005).

The final type⁹ of population recording is called event related potential (ERP). It is a technique that measures electrical activity of groups of neurons in response to a stimulus. Brett-Green et al. (2008) examined multisensory integration in children (ages 6-13) through event related potentials. The stimuli used were auditory clicks and electrical pulses. Using the superadditive definition of integration, they found four time-windows of integration. Unfortunately, no work was done to attempt to pin down the regions or pathways of the brain involved in integration. The results do show early integration (60-80 ms) followed by multiple later windows of integration (Brett-Green et al., 2008). Molholm et al. (2002) studied the timing and topography of audio-visual integration in humans using high density event related potentials. (Integration here is defined as superadditive.) They had individuals perform a simple reaction task with audio, visual and audio-visual stimuli. They found integration as early as 46 ms over the right parieto-occipital scalp (essentially early visual areas). This integration occurred simultaneously with the onset of processing in the visual areas. This suggests that either the audio information travels up to higher association areas and back down to visual areas much faster than previously thought possible or there are direct routes from auditory areas to visual areas. Molholm et al. (2002) also found later onset integration effects possibly suggesting integration in higher brain areas. One suggested site of integration was the motor cortex, which would represent the multisensory motor conversion necessary for an individual to react (Molholm, 2002). Foxe et al. (2000) performed a similar experiment using an audio-somatosensory task. They found integration 50 ms over stimulation in regions consistent with the somatosensory cortices in the postcentral gyrus (Foxe et al., 2000).

When looking at the timing data from ERPs, it is important to keep in mind the timeline of activation of brain areas. The estimates of earliest possible activation times (the fastest a signal can reach a particular region) are derived from intracranial recordings in macaque monkeys. Comparisons of monkey and human timing (over a variety of visual, auditory and somatosensory response measures) show a 3:5 monkey to human timing ratio. Average earliest activation time for V1 in monkeys is 25-35 ms, and using the 3:5 ratio, implies 41-58 ms for humans. Using the assumption that feedforward and feedback signals travel at the same speed and the 3:5 ratio, 67 ms is the estimate for how long a signal takes to start in primary auditory regions, travel up the association areas like STP, and travel back down to primary visual areas (Foxe & Schroeder, 2005). In particular, the ERP studies by Molholm et al. (2002) and Foxe et al. (2002) show integration too early to be due to feedback. This suggests there is a more direct route being used, possibly the direct connections observed through tracing studies reviewed earlier.

Overall, there are two conclusions from the cortical studies: there is multisensory integration at a low level, and there is multisensory integration at higher association areas. None of the evidence so far is conclusive. Anatomical studies do not look at functional activity of connections. Single-cell and multi-cell recordings look at a limited number of neurons. Population techniques blur the lines between populations, but give a birds-eye perspective of integration. Still, the evidence taken together suggests multisensory integration occurs at both a low level and a high level.

⁹ This is the final kind reviewed here. Others exist and have been used to study multisensory integration. Most notable, functional magnetic resonance imaging (fMRI) has been used. Although it has higher spatial resolution than some of the techniques reviewed here, the fMRI studies do not add any new conclusions; they merely are another method to confirm what the studies reviewed here show.

The cortical data collected so far does not produce a comprehensive theory of multisensory integration with testable predictions. There are three main ways to interpret the cortical data according to Driver and Noeselt (2008). First, one could say that all regions in the brain are multisensory (i.e. unisensory no longer applies because of the interactions of low level sensory areas). This goes against the ideas of specialization of function in the brain and is inconsistent with some evidence. Second, one could say there are new multimodal areas that were previously unknown due to technological limitations. This account is an ad-hoc revision of the traditional views of multisensory integration. The final interpretation hinges on the role of feedback. The authors favor the view that the new integration effects that have been uncovered are not due to just discovering new multimodal areas, but are due to feedback architecture in the sensory regions. One final possibility that Driver and Noeselt fail to mention is that a mixture of interactions in low level sensory areas and feedback from higher cortical regions could be responsible for the evidence that we see in the cortex. Computational model can help distinguish which of these theories are plausible.

4. Computational Models

The literature in the multisensory integration gives an experimental backing upon which to base my model. The literature alone may suggest particular mechanisms are at work, but computational modeling can test those proposals. There are four main computational models of multisensory integration in the field currently. Because integration in the SC has been much studied more than the cortex, all are aimed at capturing the mechanism underlying SC multisensory integration. Each model approached the problem of multisensory integration in a different way, but some overarching themes emerge when the models are examined collectively (See Figure 6).

4.1 Rowland, Stanford and Stein

One model was created by Rowland, Stanford and Stein (2007) (The Rowland model; See Figure 7). They modeled integration in the cat SC and focused on explaining how ablating/disabling the descending afferents to the SC eliminates multisensory integration while leaving unisensory responsiveness intact. The ablation evidence shows that a simple summation of the ascending and descending information cannot explain multisensory integration. They posit that the descending connections must have a very specific modulatory role. In their model, descending connections innervate a single dendritic compartment. Ascending connections innervate different dendritic compartments. Both pathways send afferents to interneurons that synapse on the multisensory output neuron. Both descending and ascending carry two pieces of information: one auditory and one visual. By coupling the descending projections, Rowland, Stanford and Stein (2007) force the descending projections to have a synergetic, modulatory role.

The model replicates multisensory enhancement, superadditivity, inverse effectiveness and the model replicates the cortical deactivation evidence. Also, the model matches how multisensory enhancement unfolds over time (Rowland, Stanford & Stein, 2007). The model was also tested for within-modal stimuli (multiple stimuli from the same modality). Real within-modal data was recorded from the SC with single-cell recording techniques. The model was compared to three other models that have been used for within-modal stimuli. The other three models were: additive (the response to within-modal should be equal to sum of the responses to each stimuli); averaging (the response should be equal to the average of the responses to each stimuli); maximum operator (the response should be equal to the strongest of the individual

stimuli responses). The Rowland model fits the real data most accurately, outperforming all of the other models (Alvarado et al., 2008).

The Rowland model accurately anticipates the response of SC neurons to multisensory and unisensory information. The model stresses both the importance of both the modulatory role of descending connections and the influence of inhibition. The descending connections have a synergetic relationship that is dependent on multisensory cues. Inhibition keeps the model within physiological constraints. One unanswered question is: what should the output from the cortical region look like at any particular time step? The model sets both the input from the senses and the descending connections from the cortex. The scientific literature has not uncovered what information is contained in the descending projects. The assumption that Rowland and colleagues' make is that the input to the ascending and descending connections is the same. This may simplify the situation, but there is a question as to whether this is valid.

4.2 Ursino, Cuppini, Magosso, Serino and Pellegrino

Ursino, Cuppini, Magosso, Serino and Pellegrino (2009) modeled multisensory integration in the SC using groups of neurons in layers rather than just using a single neuron. Their model contained three layers of neurons: one visual layer, one auditory layer and one multisensory layer (representing the SC) (The Ursino Model; See Figure 8). Each layer consisted of a grid of 40 x 40 neurons. The visual layer was directed stimulated by visual targets in the world while the auditory layer was directly stimulated by auditory targets. The multisensory layer received inputs from the unisensory layers (and sent a weak feedback connection to the unisensory layers).

Each neuron responded to a particular piece of the world (a receptive field) that overlapped with neighbors' receptive fields. (Figure 9 illustrates the concept of overlapping receptive fields with blue receptive fields for each neuron.) If there is a visual stimuli at (20,20), then the neuron at (20,20) and its neighbors (20,21),(21,20),etc. will be activated. They also added an inhibitory response within layers. Each neuron (regardless of layer) has excitatory connections to a few, nearby neighbors and has inhibitory connections to a large portion of distant neighbors. This distribution of connections is called a Mexican hat distribution (See Figure 10).

In the model, a stimulus is presented to the network (visual, auditory or both) at a particular location in the 40 x 40 grid. The unisensory layers are activated according to the Mexican hat distributions and a sigmoidal activation function. This function takes the inputs to a neuron and converts them to an output value from 0-1 (See Figure 12 for an example curve). The curvature of the function makes superadditive multisensory enhancement (multisensory response is greater than the sum of the unisensory responses) possible. If the function were linear, the multisensory response would equal to, but never greater than, the sum of the unisensory responses.

Once activities are generated using the Mexican hat and sigmoidal function, the unisensory layers send their outputs to the multisensory layer (See Figure 11). Each neuron in a unisensory grid feeds up to the neuron in the same position in the multisensory layer (i.e. visual neuron (0,5) and auditory neuron (0,5) are the only neurons that feed up to multisensory neuron (0,5)). Each multisensory neuron sends small feedback connections to its unisensory neurons.

The model's activity unfolds over time. First, the world stimulates the unisensory layers using the Mexican hat distribution of weights. Next, that activity propagates up to the multisensory layer via the connections from unisensory neurons to multisensory neurons in

conjunction with the multisensory layer's Mexican hat distribution of weights. At the same time, the world stimulates the unisensory layers again. This process continues and eventually the network reaches a steady state of activity.

The Ursino model fits multiple pieces of evidence from the literature. First, the model shows multisensory enhancement. Multisensory response of the network is greater than the response of unisensory response of the network. In particular, low-level unisensory inputs generate superadditive multisensory effects (i.e. the response of the network is greater than the sum of the individual responses of the network to the component unisensory inputs). It is also noteworthy that the settling time (the time it takes a network to reach steady state) is lower for multisensory inputs. This matches the behavioral data of multisensory reaction times being faster than unisensory responses. Second, the model shows inverse effectiveness (the stronger the unisensory inputs, the weaker the multisensory enhancement). Finally, the model shows cross-modal and within-modal suppression. If two stimuli are presented in non-overlapping receptive fields, one stimuli works to silence the other stimuli. If the inhibitory response comes from neurons within the same layer (two visual stimuli- one at (0,0) and one at (10,10)), this is within-modal suppression. If the inhibitory response from neurons within another layer (one visual stimuli at (0,0) and one auditory stimuli at (10,10)), this is cross-modal suppression. The magnitude of suppression observed fits with biological data.

The Ursino model is one of the most comprehensive models of multisensory integration in the SC. The model stresses the importance of a nonlinear sigmoidal activation function and the importance of a Mexican hat distribution of weights. The model does still have its disadvantages. It neglects to incorporate descending, modulatory projections from cortical areas to the SC. Also, it only uses a single set of parameters amounting to a computational case study. Some of the variables were set to existing biological data, but others were set according to logical arguments. Finally, the model is all hard-wired, so the model sheds no insight onto how multisensory integration could be developed by our brains.

4.3 Anastasio and Patton

Anastasio and Patton (2003) created a model of the superior colliculus in humans (here after abbreviated as the AP model, Figure 13). They wanted to show the modulatory role of descending projects and explain why there is a mixture of unisensory and multisensory neurons in the SC. The model is similar to the structure of Rowland and collaborators' model, but Anastasio and Patton use a slightly more abstracted neuronal model. They model a 10 x 10 grid of neurons in the deep superior colliculus layer that corresponds to a tiny, overlapping receptive field. The grid receives primary inputs from the visual, auditory and somatosensory systems. The primary inputs are modulated by visual, auditory and somatosensory projections from the parietal cortex. The big difference between the two models is that the connections in the AP model are trained instead of hard-coded.

Because they were modeling a tiny receptive field, Anastasio and Patton were able to simplify the inputs to their model. Whether inputs are stimulated or not depends on probabilities. First, they set the probability of a target appearing in the field. Then they set the probabilities that a target that appears is of a particular modality or of a combination of modalities. Each time stimulation is needed, a target is selected according to the probabilities. The modality of the target is then used to stimulate both the primary and modulatory pathways.

The connections in this model are trained through a two phase learning procedure. The first phase (named the Hebb phase) runs a standard unsupervised self-organizing map (SOM)¹⁰ on the primary inputs. The second phase (named the Hebb-anti-Hebb phase) runs a novel rule on the modulatory inputs. The activation of a neuron is found by passing the weight sum of the inputs through a sigmoidal function.

In the Hebbian step, a target is selected according to the probabilities set out earlier. This only activates the primary inputs because the modulatory inputs remain silent during this step. Now the SOM is applied to the network. The activity of the SC neurons is calculated. A winner (the neuron with the highest activity) is selected, and the neighborhood (neurons that are topographically nearby) of that neuron is found. The weight update works by reinforcing the winner the most and reinforcing the neighborhood based on how far away a neighbor is away from the winner. (The farther away the neighbor is the less of an update it receives.) Figure 14 provides an example of the training regiment with black as the winner. The shades of gray form the neighborhood, and the blue represents the neurons not in the neighborhood. The darker the shade of gray the stronger the reinforcement will be. This winner-neighborhood reinforcement is repeated for a specified number of steps with decreasing reinforcement for each subsequent step. In other words, the learning rate decreases after each iteration of the reinforcement.

This first learning phase forces the SC neurons to extract information from the primary inputs. Anatasio and Patton (2003) measured information gain during training and found the largest information gain comes from this step. They also found that by changing the probabilities controlling the target and the thresholds of the neurons, any mixture of multisensory and unisensory responsive neurons can be generated.

In the Hebb-anti-Hebb phase, the modulatory weights are trained. A target is selected based on the probabilities earlier defined. Both primary and modulatory inputs are activated based on the target. No training occurs on the primary weights. The objective of this stage is to fulfill two constraints: modality-matching and cross-modality. Modality-matching means that the inputs to a SC neuron should only be modulated by a modality that matches one of the primary inputs to the SC neuron. Cross-modality means that primary inputs should never be modulated by a modulatory connection of the same modality. (See Figure 13b.) The neuron is only allowed to have modulatory connections from auditory and visual because of the modality-matching constraint. Because of the cross-modality constraint, the visual primary input can only be modified by the auditory modulation projection, and the auditory primary input can only be modified by the visual modulation projection. The literature suggests that this is the parietal descending projection's pattern of innervation in the SC.

There are three rules that form the Hebb-anti-Hebb training.

1. If a SC neuron and a modulatory input are both active, then decrease the modulation of the active primary inputs by B and increase the modulation of the inactive primary inputs by B.
2. If a SC neuron is not active, but a modulatory input is active, then decrease the modulation of all primary inputs by 2B.

¹⁰ Hebb's principle is that neurons which fire together wire together. Essentially, anytime two interconnected neurons fire synchronously, their connection is strengthened (Bienenstock, Cooper & Munro, 1982). The SOM the AP model uses is merely an extension of the basic Hebbian principle. Imagine two neurons (A and B) in a population of neurons. If A has a connection to B and A fires, Hebbian rules dictate that the connection from A to B must be strengthened. SOM rules say that in addition to the connection from A to B being strengthened, connections from B to neurons that are close to A should be strengthened (Kohonen & Hari, 1999).

3. If a modulatory input is inactive, do nothing

Rule 1 forms the cross-modality constraint. Rule 2 is the modality-matching constraint. Rule 3 just specifies what to do if a modulatory input does not fire. The only additional caveat is no weights are allowed to be negative; if a weight drops to a negative value, it is immediately changed to the value 0.

This second stage produces multisensory enhancement through its unimodal, modulatory projections. Superadditive multisensory enhancement only is found after the modulatory connections have been trained. If the modulatory connections are cut, some subadditive enhancement can be observed, but the superadditive enhancement is eliminated while still maintaining unisensory responsiveness, similar to what is seen in the experimental data for SC.

Interestingly, the model fits well with the data from the development of the SC in cats mentioned earlier. Weak multisensory response is found in cats until a critical point in which adult-like multisensory response appears. It is possible that a two-stage learning procedure like the AP model is at work. Also, increasing the amount of noise in the inputs increases the percentage of multisensory neurons. This lines up somewhat with the principle of inverse effectiveness; the noisier the data, the more integration can reduce variance.

This model has some of the same drawbacks as Rowland and colleagues' model. The AP model does not account for multisensory depression and does not adequately address the setup of the action of downward projections. Additionally, this model does not allow for any inhibitory connections which goes against the known importance of inhibition in the brain. Despite the drawbacks of the models, they both show the modulatory pattern of descending afferents is essential to accounting for multisensory integration in the SC.

4.4 Martin, Meredith and Ahmad

Martin, Meredith and Ahmad (2003) built a model of multisensory enhancement (MSE) using the principles of traditional SOMs in conjunction with a sigmoidal transfer function (a nonlinear neuronal activation function) (hereafter referred to as the MMA model). Their work is a simplification of Anastasio and Patton's 2003 model. Whereas AP required a two-stage learning procedure (Hebbian SOM stage followed by Anti-Hebb-Hebb stage), MMA showed that MSE can be generated without the second stage. Their goal was to create a model of MSE that was not specific to the SC.

The setup of their model is very similar to the AP model. The model consists of a number of modalities or senses. Each sense is a collection of 20 sensory receptors, so the activity of a particular sense can range from 0-20. These 20 sensory receptors are treated as a unified entity that sends projections to the SOM. The training procedure rearranges the connections from modalities in order to extract information from the senses in the same way that AP's first stage worked (See Figure 16).

In each step of training, a modality string is generated. A modality string is a bitstring that represents whether a sense is active or not. A 1 in a bitstring means the corresponding sense is actively driven while a 0 in a bitstring means that the corresponding sense fires at noise level. In a three modality setup, the bitstring 010 means that modality 2 is the only active modality; modality 1 and 3 are set to have noise level activity (See Figure 14 for all bitstrings for a 3 modality system). The values for activity and noise level are calculated based on probability functions that are set as parameters of the model. MMA set their driven level to a binomial distribution centered on 12 and their noise level to a binomial distribution centered on 2.

MMA found that if they train a SOM on a 2 or 3 modality system in the same manner as AP, the SOM shows robust MSE. The map is specialized to have a unisensory region for each modality. The MSE shows up in the regions in-between the unisensory areas. See Figure 17 for a visual representation of their results.

The MMA model is surely a simplification compared to AP and Ursino models' system level approach. It lacks the complexity of inhibition, MSD, time dynamics and descending/modulatory connections. MMA also only used a single set of parameters to test their model meaning that the work results to a computational case study. Even with all these missing elements, MMA managed to find robust MSE using a simple, biologically inspired technique.

4.4 Context

Each model has its assumptions and insights. Rowland et al. (2007) stress the importance of hierarchy of connections and inhibition, but their model limited its scope to the neuronal level. Ursino et al. (2009) created a complex hierarchy with Mexican hat distributions of weights, but they failed to incorporate the role of descending connections. The AP model focused on the hierarchy of ascending and descending connections, but failed to explain how MSD occurs. The MMA model made the most simplifications by far, but the end results was a simple, biological rule to generate MSE. Although their model was intended to explain the whole story of MSE, I believe they engineered a possible fundamental building block of multisensory integration. Although the MMA model relied on the SC literature, the model was intended as an abstraction that could be applied to any multisensory integration- not just that limited to the SC.

5. My Work

I set out to try to bring the computational modeling of the SC to a more cortical setting. The first step to achieve that was to reassemble the MMA model. I needed a deep understanding of MMA model's building blocks before I could use it to construct anything more complex. Given the MMA model's case study approach to parameter setting, I needed to understand exactly how this model works under many conditions. The second step was to use this understanding of the MMA model's to make extensions and start bridging the gap between this simplified model and the more comprehensive models. The final step was to bridge the gap between the models of SC integration and the literature on cortical multisensory integration.

Before jumping into my work, I will walk through MMA's model in more detail. A visual summary of the model can be seen in Figure 15.

5.1 The MMA model in Detail

There are three modalities. A modality is a collection of 0-20 sensory neurons. The modalities are wired to the SOM randomly to start but are trained on a standard SOM procedure. Each step in the training procedure involves the following:

- 1) Selection of a random modality bitstring
- 2) Activation of the modalities based on the bitstring
- 3) Propagation of activity from modalities to the SOM neurons
- 4) Activation of the SOM neurons
- 5) Weight update
- 6) Weight normalization

In the first step, the inputs are bitstrings as mentioned earlier (000,001,010,etc.). 0 means a modality fires at noise level, while 1 means a modality fires at an active level. Each of the

bitstrings of length 3 is equally likely to be generated. This leads to multisensory stimuli being presented half of the time. One of the bitstrings is selected at random.

Second, the bitstring is translated into the activity level for each modality. Because the modalities are just clusters of 20 sensory neurons, the activity level of a modality is equal to the number of sensory neurons that are firing within that modality. If a modality is driven, each neuron has a 60% chance of firing. If a modality is spontaneously firing, each neuron has a 10% chance of firing. Translating the bitstring to activity levels amounts to sampling from a binomial distribution centered on 12 (for 60% of 20 neurons) or 2 (for 10% of 20 neurons).

Third, the activity of each modality (a number between 0 and 20) is passed through the weighted connections from each modality to each SOM neuron. Each SOM neuron then has a weighted input sum which amounts to how much the input modalities are stimulating the neuron. In the fourth step, these input sums are translated to activity levels by passing them into a sigmoidal function (See Figure 18). This function takes a value on the scale of 0-60 and puts it in the range of 0-1. This threshold function is supported by the literature (and will be discussed in more detail in the next section).

In the fifth step, the weights are updated. The winning neuron (the neuron with the highest activity level) is calculated. Each unit is updated based on proximity to the winning unit. Those that are farther away are not allowed to change as much, whereas those that are close are allowed to change more. Each unit's weight changes to better align with inputs. Imagine a SOM neuron connected to sense 1 with activity a_1 and sense 2 with activity a_2 . When the SOM unit's connections are allowed to update, the connection from sense 1 to the SOM neuron tries to move its value closer to a_1 and the connection from sense 2 to the SOM neuron tries to move its value closer to a_2 . If $a_1 = 0$ and $a_2 = 20$, the connection from sense 1 will be decreased (the weight will move toward a value of 0) while the connection from sense 2 will be increased (the weight will move toward a value of 20). After updating, the weights to a unit are normalized by unit length. In the final step, weights are normalized by vector length. The weights to each SOM neuron can be represented as a vector (w_1, w_2, w_3) . w_x is the value of the connection from sense x to the SOM neuron. The length of this vector is the square root of the sum of the squares of each element in the vector. If each of the weights is 1, then the length is the square root of 3. Normalizing by the vector length means that each element in the vector gets divided by the vector length. This preserves the direction of the initial vector, but changes the distance it travels to 1.

This makes intuitive sense in the two dimensional case. Imagine a weight vector (4,2) that represents connections from auditory and visual inputs to a SOM unit. The length of this vector is the square root of the sum of the squares of the elements (e.g. square root of (4^2+2^2)). This length is the length of the hypotenuse formed from the vector. The normalization procedure divides each element in the vector by the length of the vector. In a 2-D situation, this amounts to dividing each side of the triangle by the hypotenuse. The hypotenuse is about 4.47 for (4,2), so the normalized vector is about (0.89,0.45). This point lies on the unit circle (a circle centered on the origin with a radius of 1). Now imagine many 2-D vectors: (4,2), (10,1), (223,104), (22,5). How could you compare the x and y values without magnitude being a factor? Normalization by unit length places all of these vectors on the unit circle without changing the direction. The vectors can all be scaled so that direction can be compared without magnitude being a factor (See Figure 18).

This six step training procedure is repeated for 5000 iterations after which the networks weights are considered to have been trained. The next step is to analyze the results. MMA chose to examine the activity levels and MSE.

To analyze activity levels, MMA would choose a bitstring and rerun their model through steps 2-4 of their training procedure (up through the point of activation of SOM neurons). They repeated this process 1000 times for each bitstring and calculated an average activity of each SOM neuron. You can see the results in Figure 17. The bitstrings above each box indicate what bitstring was used to generate that particular box. Each box is a 10 x 10 grid representing the 10 x 10 SOM. The colors indicate the level of activity from 0 (blue) to 1 (red). As you can see from the graphs, there are circles of unisensory responsiveness in the SOM. These can be found in the corners of the SOM. You can also see distinct bimodal areas that are located in-between the unisensory corners. If the lower right corner is sense 1's area (responds strongly to 100) and the lower left corner is sense 2's area (responds strongly to 010), then the area that responds strongly to the bimodal case (110) will be in-between the lower left and lower right corner. The activity levels for 111 (all active) and 000 (not active) can also be observed in the last row.

These distributions of unisensory and multisensory areas are suggestive of multisensory enhancement. The amount of MSE (multisensory enhancement) can be calculated using the multisensory index (a common multisensory metric). The multisensory index is: $(MS - US_{max})/US_{max}$ where MS is the multisensory response and US_{max} is the maximum of the component unisensory responses. The metric shows how much greater the multisensory response is relative to the maximum unisensory response. MMA used the average activity levels they calculated through 1000 presentations of bitstrings to calculate the MSI. These MSI results can be found in the third and fourth row of their figure; they are also color coded from blue (0% MSE) to red (250% MSE). Bimodal MSE appears as a line of enhancement in-between the two unisensory regions involved while the trimodal MSE appears as a triangle in-between all the unisensory areas. By training a network on this six step procedure, the MMA model shows robust regions of multisensory enhancement.

5.1 Replication

My first goal was to recreate the MMA model and alter parameters to gain a deep understanding of how the SOM and MSE work. After some communication with Martin, I was able to replicate their model. Visualization was the next step. Regardless of what the model does with vectors and numbers, how can you see what is happening? I choose to look at the weight distributions, activity and two measures of MSE.

The weights to each SOM unit were trained and normalized when running the MMA model. The result is that each SOM unit has a weight vector where each weight in the vector ranges from zero to one. I choose to visualize each vector using colors. The three senses are assigned a different color (red, green or blue). Each SOM unit is a square with three rectangles. Each rectangle's color represents the weight of the connection from a modality to the SOM unit. Hence, a SOM unit that has a strong weighted connection from the visual sense would have a bright red rectangle, whereas a SOM unit with a very weak connection from visual would have a dull red or black rectangle. The end result of this analysis is a grid of colors that represents how much each SOM unit pays attention to each of the modalities (as can be seen in Figure 20).

The upper left hand corner of Figure 20 has very strong connections to the third modality and very weak connections to other senses. The lower left hand corner has very strong connections to the first modality and very weak connections to other senses. The right side of

the grid is specialized for the second modality's inputs. The regions in-between these unisensory specialized areas form a gradient. As you move from the upper left to the lower left, you see connections from sense 3 become progressively weak as connections from sense 1 become progressively stronger. The middle ground between the two areas is about even in weights. This gradient trend is a topographical map, i.e. the spatial position has a meaning. The SOM units near each other respond to similar inputs. Moving in a particular direction on the SOM grid changes the weights in a particular direction (moving down increases the visual weights, moving right increases the auditory weights, etc.).

Activity was analyzed by stimulating a trained network with a particular stimuli (001, 010, 100, etc.) 1000 times. An average activity of each SOM unit was thus obtained. This can be visualized by color-coding the activity of average SOM activity. As can be seen in Figure 21, warmer colors represent activity closer to one while cooler colors represent activity closer to zero. These activity graphs yield essentially the same information as the weight distributions, but in a more readable format. The corners/sides of the SOM are specialized to handle a particular sense. Bimodal inputs show strong activity in-between the unisensory areas. Complete noisy input (000) shows very little activation on average. Complete signal input (111) shows very high activation in the center of the SOM grid. Again, these trends match the graphs in the MMA paper.

In analyzing MSE, I used two similar, but distinct metrics. First, I followed the average multisensory index procedure laid out by in the MMA paper: find average activity levels after 1000 presentations of a stimulus and calculate the multisensory index. My results line up with the MMA paper; MSE is in the range of 0-300% that shows up in-between the unisensory regions (See Figure 22). By calculating MSE of average levels of activity, variability in the data is prematurely removed. There is no a priori reason why the MSE of the average level of activity represents how much MSE is happening on average (See Figure 23 for an example).

My second multisensory metric was only different in when the average was taken. In each of the 1000 presentations, I calculated the MSE for the individual data points and then averaged the 1000 MSE values (as shown in Figure 23). The same general trend showed up in with this second metric, but the magnitude changed from 0-300 all the way to 0-700 (See Figure 23). This is more than double the amount reported by the MMA paper as fitting the biological data. My MSE metric may better approximate the values from single cell recordings, whereas the MMA protocol may better approximate the MSE values observed using population recording techniques (like fMRI, etc.). In any case, we must be very careful when fitting MSE results from a model to biological results.

These trends change slightly based on random seed¹¹ (See Figure 24). The position of the unisensory areas (the SOM units that pay close attention to a single modality) may switch. Auditory areas could be in the upper left or the lower right, etc., but every map forms an auditory area, a visual area and a somatosensory area. Every map also forms the same topographical relationship: moving toward a unisensory area increases the weight from that modality. Given this dynamic, you can always expect the trimodal area (the area that pays attention to each sense equally) to lie in the center of the SOM grid because that is the area that is always in-between the corners.

¹¹ Random functions work in Python by reading the next number from a long list of pseudorandom numbers. Random seeds set the starting point of where Python will look in the list. By running the same program with the same random seed, you can exactly reproduce the same results.

5.2 Unmodified MMA

Once I successfully replicated the results from the MMA paper, I began a series of many experiments to tease apart the model. There are four main categories of experiments that I performed on the MMA model solely by adjusting parameters: SOM formation; activity probabilities; noise effects; and sigmoid alterations.

5.2.1 SOM Formation

The first set of experiments was aimed at altering what seemed to be the biggest SOM formation parameters: the width and SOM size parameters of the model. These are complementary parameters. Width changes the neighborhood function of the weight training. A larger width means that more neurons near to the winning neuron are strongly reinforced (i.e. the neighborhood is larger). A smaller width means the opposite. MMA used 1.0 as their width. The SOM size defines how many neurons are in the self-organizing map. MMA used a 10 x 10 grid (100 neurons). I performed a series of experiments in which I either held everything but width or SOM constant. Looking at the trend over 3, 10, 30 and 100 size grids (with a width of 1), the SOM quickly loses its ability to hold together (See Figure 26). In the 30x30 grid, the areas of unisensory responsiveness have expanded. In the 10x10 grid, unisensory areas were just circles, but now they have extended a tail of activity. In the 100x100 grid, the map has been splintered completely. Pockets of specialization into a particular modality can be found, but the map has no overriding pattern. These splintered maps can be “saved” by altering the width parameter. By starting with a 50 x 50 grid with a width of 1.0, I was able to reestablish unisensory areas (See Figure 27). It’s worthy to note that the width can get too big as can be seen with the 50x50 grid that has a width of 30. Width and SOM size must be kept in the right proportions. The neighborhood as defined by the width must be able to grab a single chunk of the SOM without grabbing everything. In other words, the SOM algorithm is scalable if width and size are balanced. This means that SOMs are a promising area of study when dealing with a brain that has billions of neurons.

5.2.2 Activity Probabilities

I coded MMA’s model so that I could alter the driven and spontaneous probabilities for each sense independently. I began systematically changing the probabilities (making different senses “stronger” and others “weaker”). First, I altered the driven probabilities for all the senses while leaving the spontaneous probabilities at 0.1. I used 0.5, 0.6 (0.6 was the MMA value), 0.7 and 0.8 for the driven levels (See Figure 28). Every value changed the magnitude of MSE and activities, but the same trend observed with the MMA initial parameters holds. Using 0.5 for the driven probability increased MSE values for the trimodal case (from 300 to 400 with MMA’s metric and from 700 to 900 with my metric). As the value increased to 0.7 and 0.8, the responses of the network to unisensory inputs grew relative to the response to multisensory inputs. The result was a large decrease in MSE values observed (by about ½). Another set of experiments where I boosted the spontaneous probability levels to 0.3 while keeping driven probabilities constant at 0.6 showed a large increase in the unisensory areas. This resulted in no MSE being detected with the MMA metric and very little bimodal MSE observed with my metric. Alterations of spontaneous and driven levels independently for each sense showed the same trends. These results (supported by later explanation in the sigmoid section) show that

MSE magnitudes can be manipulated very easily, so matching magnitude to biology must be done very carefully.

5.2.3 Noise Effects

The next set of experiments was aimed at understanding what noise is in the MMA model. Typically, we think of noise as the enemy. We try to limit noise as much as possible in order to boost signal. Increasing the noise through changing the spontaneous and driven probabilities (in the prior experiment) did not cause the network to become random and noisy. Instead it had the puzzling result of expanding the unisensory areas. I decided to probe deeper by altering the bitstring probabilities. I changed the probabilities so that only unisensory stimuli were used (001,010,100). As you can see in Figure 29, this does not change the map formation greatly. Next, I changed the probabilities so that only the no stimuli condition was allowed (000). Again, this strangely did not affect map formation. The biggest change I could produce was with only allowing trimodal inputs (111). Even in this case, the essentials of the map have not changed (unisensory and multisensory regions). All of these experiments altered the gradients formed, but none broke the general pattern observed.

I suspected noise was the answer as to why the MMA model was so robust to changes in the bitstrings. I modified the code slightly to eliminate both the spontaneous firing level and set the active firing level to a constant (12). The result can be seen in Figure 30. The maps are fractured and very discretely divided up. Noise was essential to creating the smooth maps in MMA's paper.

In order to explain these results, it is useful to reexamine self-organizing maps from a mathematical perspective. Self-organizing maps work by categorizing/organizing input, as you can see in Figure 31 (reproduced from Kohonen 1990). The network he set up is trained on arbitrary items labeled A through Z plus the numbers 1 through 6. Each item corresponds to a unique value among the attributes (a_1 through a_5). A corresponds to $a_1=1$, $a_2=0$, $a_3=0$, $a_4=0$ and $a_5=0$, represented by the vector (1,0,0,0,0). The self-organizing map is trained on presentations of random attribute vectors that correspond to one of the items. When an item is presented, the winning SOM unit is the one that best represents that particular item. Neighbors of the winner store similar items.

If we look at the items again, we see that A, B, C, D and E are all very similar. A (1,0,0,0,0) is one value different from B (2,0,0,0,0) which is one value different from C (3,0,0,0,0), etc. When we look at the map, we see that A is next to B, which is next to C, etc. When A was presented during training, the upper left unit was the winner. It moved its weight vector to match A. Its neighbors also move their weight vectors to match A. The same happened when B, C, etc. were presented. Because the vectors for the items were similar, this forced the map to keep A, B, C, D and E clustered together. The SOM organized all the items based on vector similarities.

To illustrate the effects of noise, imagine training a SOM to produce a color map. Each example the SOM is trained on is a color (a vector with red, green and blue values between 0 and 255). White would be represented by the vector (255,255,255), a bright red would be (255,0,0), etc. If the network is trained on three colors bright red, bright green and bright blue, the resulting map would have a jagged gradient. (255,0,0), (0,255,0) and (0,0,255) would be spatially arranged, but because their values are so disparate, the map cannot blend the cases smoothly. Any particular color value is either 255 or 0. Units in the map will be designated to

respond to either 255 or 0. The only gradients formed would be by the neighborhood function (which allows neurons closer to the winning unit to be updated more than those farther away).

Now imagine the same color map, but with three more colors in our training set: yellow (255,255,0), purple (255,0,255) and aqua (0,255,255). Each item in our training set is more closely related to some colors than others. Red (255,0,0) is more similar to Yellow (255,255,0) than to Green (0,255,0). Now our color map has more cases and hence a fuller map.

The principle of why noise creates a smooth gradient is based on the fact that more training cases leads to a fuller map. Take our color map again. Instead of explicitly adding colors, let us add random perturbations to our colors when we present them for training. When we intend to present Red (255,0,0), we may present (250,1,2) or (255,10,0) and so forth. Our map tries to organize each of these cases just like it organized Red, Green and Blue in the first example. The network treats Red (255,0,0) and slightly Off Red (250,1,2) as two separate colors that it must organize, but because their values are similar, they are organized nearby each other.

Now our map will appear to map colors with smooth transitions. Under the hood, the map really just has more cases in its repertoire, many of which are tightly clustered. All the Red hues are next to each other. All the Red colors that have a little green added to them are closer to the area of the map where the Green hues are located. This distribution gives the appearance of a full, smooth mapping.

The same principle applies to the SOMs used in the MMA model. As you can see in Figure 30, turning off noise causes jagged map formation. We tend to think of noise as an enemy with which our brains constantly grapple. The role of noise in the SOMs may say otherwise: our brains are hardwired to take advantage of noise in our environment. A certain amount of noise appears helps to form complete self-organizing maps. The advantage of a complete SOM is that there are fewer unexpected cases. Imagine our initial red, green and blue SOM being given a yellow light. The SOM would act in an unexpected manner because of the lack of a gradient between colors. Noisy inputs generate more cases for a map to organize and better prepare for. Noisy maps are prepared for noisy inputs.

5.2.4 Sigmoidal Alterations

The final set of experiments that I performed was aimed at understanding the sigmoid function which is the basis for the MSE observed. The sigmoidal activation function takes an input sum and converts it to a number between zero and one. The input sum is the sum of the weighted activity levels of the modalities. If an activity vector is (12,12, 2) and the corresponding weight vector is (0.71,0.71,0) , the input sum is $12 * 0.71 + 12 * 0.71 + 2 * 0 = 17.04$. The sigmoidal function (as seen in the Figure 18) puts 17.04 into the range from zero to one. The curve is designed to simulate neuronal thresholds. Neurons fire in an all-or-nothing process. If the input sum to a neuron is greater than a particular value (the threshold), the neuron will fire. If the sum does not exceed threshold, the neuron does not fire. The sigmoidal's s-shape approximates this firing pattern¹².

For input values from 1 to about 6, the transfer function spits out values very close to 0 (tiny slope). From values of 7 to about 13, each additional input yields increasing gains in activity (i.e. the slope is positive and increasing). From the values of about 13 through about 19, each additional input yields decreasing gains in activity (i.e. the slope is positive but decreasing). The rest of the curve (values greater than 19) yields values very close to 1 (tiny slope).

¹² http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#A more complicated neuron

The nonlinearities of this curve make MSE possible. If a neuron is stimulated with an input of 10 in the case 001 and in the case 010, that yields an activity of about .316. The bimodal case (011) would have an input sum of 20, which yields an activity of .986. A calculation of the MSI would be $(.986-.316) / .316 * 100 = 212\%$ MSE. The same calculation with unimodal inputs of 7 yields a MSE of about 700% (See Figure 32 for a visual example). If we were to perform the same MSI calculate on the input levels, we would get $(20-10)/10 * 100 = 100\%$ MSE. 100% is the maximum MSE you can obtain using linear summation. This does not fit with the biological data.

The sigmoid comes with two parameters: tonic bias and sensitivity. Tonic bias controls the horizontal shift of the curve. A higher tonic bias shifts the curve right and a lower tonic bias shifts the curve left. A value of 12 for the tonic bias says that it takes an input sum of 12 to get 0.5 activity. See Figure 33 for a comparison of sigmoidal curves with biases of 11 and 17. The second parameter, sensitivity controls the slope of the central part of the curve (Ursino et al., 2009). Increasing sensitivity enhances the non-linearities and decreasing sensitivities moves the curve closer to a straight line (See Figure 34). MMA sets the sensitivity at $1/2$, but does not specify reasoning for why that particular value. AP sets their value to $1/5$ and Ursino et al. set their value to 0.3. MMA sets their tonic bias to the maximal response a single modality can generate in a multisensory neuron¹³. This logic forces SOM neurons that are multisensory to show high levels of MSE, while unisensory neurons must show little to no MSE. Since the driven probability is set to 0.6, the inputs from modalities that are driven are 35% of the way up the curve (60% of the way to 0.5). This point lies in the area of the curve with a small slope section of the curve. Doubling this input puts the activity above the central point and into the region of decreasing, positive slope. See Figure 35 for a visual representation. The red shade shows the region of the curve where unisensory responses that generate high MSE can be found.

In experimenting neuronal response, I altered the sigmoid transfer function (adjusting the slope and shifting the curve left or right). In doing so, I realized that MSE is completely dependent on this curve. The results hinge on this crucial bit of code. In general, the slope of the function will determine the magnitudes of MSE observed and the horizontal position of the curve will determine what types of input generate MSE. As described earlier, there is an optimal region on the sigmoid for MSE. You want the unisensory stimuli to lie as far to the left of inflection as possible while having the multimodal stimuli lie as far to the right of the inflection point as possible. Decreasing the slope of this function moves the curve closer to a linear function (decreasing nonlinearities necessary for MSE) while increasing the slope of this function creates a greater jump around the point of inflection (potentially increasing MSE). Shifting the curve left brings noise level inputs closer to the point of inflection, potentially generating meaningless MSE. Shifting the curve to the right moves noise level farther away from MSE, but also moves driven levels farther away from MSE. The sensitivity of results to changes in the transfer function is to be expected.

The control over the sigmoid and the state probabilities makes it possible to show MSE in many areas within similar networks. Without some biological standard for parameter setting, MSE is an artificial, mathematical construct. The results of experiments could have little meaning (noise showing MSE, etc.).

¹³ This can be found through mathematical calculations. If we are dealing with a three modality scenario, a multisensory neuron that pays attention to every modality would have the weight vector (1,1,1) before normalization. $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ is the normalized weight vector. Since a particular modality can only output 20 at maximum, the maximum activity from a single sense is $20 * (1 / \sqrt{3})$.

5.2.5 Conclusions

By testing the original model under various conditions, I have shown some of the essentially components of the SOM based model of MSE. By adjusting size and width, I can scale maps to virtually any dimensions. This means that it may be possible for SOMs to be acting on large scale networks in the brain. My experiments with noise show that not only do maps handle noise well, but they convert it into information. Instead of trying to reduce noise, SOMs embrace it. Our brains are intrinsically noisy, so algorithms that our brain runs must be able to deal with noise. Although size and noise experiments revealed positives of SOMs, my sigmoid and state probability experiments cast some doubt on the usefulness of using SOMs alone for multisensory integration. The areas of a network that show MSE can be easily manipulated. For SOMs to be biologically plausible, MSE should line up with functional benefits of multisensory information (like in the behavioral studies covered in the literature review). With these insights in mind, I set about trying to extend the MMA model to be more biologically realistic.

5.3 Extensions of MMA

My model is an extension of the MMA model (and hence an extension of the AP work). Both MMA and AP made two large simplifications in their models. First, the modalities in the models were represented by single nodes. The idea was that their models examined MSE at the level of two neurons with overlapping receptive fields. If a sense was driven, there was a stimulus present in that sense's receptive field. The logical next step for these models is to find out what happens when these mechanisms are applied to a more complex set of modalities. What happens if a single sense consists of multiple neurons with their own receptive fields? Do two of these senses generate MSE? My first goal was to see if the same computation that worked at the level of a couple of cells could scale up to the next level in hierarchy. Both models also made the decision to leave out inhibition and hence also leave out MSD. Inhibition is essential to proper brain function and MSD cannot be generated without it. MSE and MSD appear to be complementary processes. A model that accounts for one without the other is incomplete and possibly inaccurate. My second goal was to add MSD to the MMA model.

I began working towards reproducing the MMA model, and with the help of the lead author, recreated their robust results. In addition, I coded multiple visualization procedures to color code information about the SOMs (like weight strength, MSE, etc.). Once I clearly had their model replicated (and optimized for speed using Psycho, a python library), I began to experiment with extending the model with the goal of creating more complex senses and incorporating inhibition. After many modifications of their code, I have four important experiments. The first involves a simple extension of the MMA model to 2x2 senses. The second experiment involved pushing the limits of how big senses themselves can be. The third experiment brought inhibition into the model. Finally, I attempted to add an additional layer of complexity to the senses by adding overlapping receptive fields.

5.3.1 General Extensions

Before describing the experiments, I should mention a few general changes I made to the model that apply to all these experiments. I instilled meaning into all of my inputs so that the results could be interpreted with meaning in mind. The MMA model uses arbitrary probabilities, which are great for a mathematical analysis, but since MSE has been tied to behavioral effects, I wanted the MSE to be meaningful. To add this meaning, I removed MMA's code for state

probabilities and replaced it with a virtual world. A configuration of the world is generated each iteration of training. The configurations had various properties depending on what sense I was working with. All of the experiments I will present are based on a location sensor, so each time a world configuration is generated, a random world location is chosen as the stimulus (See Figure 36). (Other experiments used different senses like red, blue and green color sensors or combinations of senses like one location sensor and one red light sensor.) The success of a mapping can be assessed by looking at how different the activity levels of the SOM are when stimulated with each possible location in the world: a measure of how well the SOM can tell each location apart.

5.3.2 MMA Model with 2x2 Senses

The first experiment involved a simple extension of the MMA to use two 2x2 senses instead of two point sources. The senses were modified location sensors: one could only sense the x position and one could only sense the y position (See Figure 37 for example activation). Each sense also had an accuracy level associated with it. Each time a location was to be sensed by a modality, a random variable was generated. If the random value was less than the accuracy level, the modality sensed the stimulus normally. Otherwise, the modality was spontaneously activated (0s for each grid in the sense). Parameters were held at the same values MMA used except for the sigmoid function. Moving to a more complicated sense required some changes. First, the tonic bias had to be changed. Although the same procedure was followed (setting the tonic bias to the maximal output of a single sense through a multisensory neuron), the definition of unisensory neuron and multisensory neuron has changed. A unisensory input refers to the configuration of one of the location senses, not to the output of any of the elements within a sense. Instead of 1 being a unisensory input like in MMA, a unisensory input was 1100, 0011, 1010 or 0101 (reading the grid from left to right and then up to down). A multisensory neuron pays attention to 11001010, 11000101, 00111010 or 00110101. A normalized weight vector can be generated on the expected activity levels. The maximal input from a unisensory is 20. Results from the experiment with this tonic bias can be seen in Figure 38. I used multisensory index as MMA measured it to analyze my networks. I found the SOM created pockets of unisensory responsiveness and multisensory responsiveness. The multisensory areas were positioned in-between the unisensory areas like in the MMA results. The MSE shown occurs between the unisensory areas also like in the MMA results. Unfortunately, the MSE observed was a magnitude greater than the MMA observed levels (6000 vs 300). I had to alter the sensitivity of the sigmoid to remedy this disparity in levels. By shifting the sigmoid further to the right, I have decreased the proportion of nonlinear sections in the curve (compare the sigmoidal curves in figure 34). To compensate, I had to stretch the curve out more by decreasing the sensitivity (to the level used by AP). By doing this, I found MSE on the level of 300 percent matching MMA's results (see Figure 39).

I also experimented with the accuracy of the senses (see Figure 40). High accuracies drive the SOM mapping of the unisensory areas closer together resulting in tighter MSE areas. Lower accuracies drive the SOM mappings farther away resulting in sloppy maps that do not distinguish locations apart very well.

The final experiment I performed was to examine what would happen if I had a single sense respond to two stimuli. Instead of the activities looking like 0011 or 1100 and 0101 or 1010, they looked like 1111. I found that two stimuli within the same modality generated

substantial within modality enhancement (See Figure 41). This result would be akin to your brain looking for a single visual target, but finding two. Instead of trying to figure out which is the real visual target, the SOM enhances activity levels. This points to an inadequacy of the model. The SOM algorithm is an organizer that treats information within a sense in the same way that it treats information about a relation between two senses; it tries to map both spatially. Even though we see MSE in some of the right places, the erroneous MSE observed within a sense makes it very unlikely that the MMA algorithm alone is responsible for multisensory integration in the brain. None of the literature points to MSE being the result of two disparate sensory stimuli. Instead, the literature shows that MSD is likely to occur; one of the stimuli should be silenced through inhibition. Without inhibition in the MMA model, a complete story of multisensory integration in the brain cannot hope to be reached.

Overall, the 2x2 sense was a success, showing robust MSE in-between unisensory areas. Unfortunately, there were two unsettling caveats: the sigmoid had to be tuned and there was MSE within a sense. These results suggest that 2x2 senses may be the limit of the MMA model without additional modifications.

5.3.3 MMA Model with Larger Senses

My next extension of MMA was to test how big I could make my gridded senses. In this series of experiments, the two location sensors were identical in setup and both sensed the entire world (not just the x or y direction). The location sensors also had an accuracy level like in the first extension. I started with 2x2 grids, and then moved to 3x3 grids and finally 4x4 grids. I set the sensitivity to AP's value like in the first extension. I set the tonic bias as specified by the MMA procedure. For the 2x2 situation, multisensory inputs look like 1000 1000 or 0100 0100, etc. This results in a maximum output of a single sense through a multisensory neuron of 15.3. The same procedure can be used to find the maximums for 3x3 and 4x4 grids. All other parameters were held at the same levels as the MMA model. The results in Figure 42 show that the sigmoidal tuning procedure holds for 2x2, but starts to lose effectiveness at 3x3 and is completely wrong for 4x4 grids. The problem is that MMA disregarded spontaneous firing in their tuning procedure. As the size of the grids increases, the amount of noise in the system increases, but the amount of the driven activity level stays the same. That is why MSE is observed all over the place in the 4x4 grid. The noise level activities have crept their way into the area of the sigmoid that generated MSE.

In a complementary experiment, I modified the bias manually to move the MSE from the noise to the driven levels. This worked to an extent (See Figure 43). While the MSEs look more close to the 2x2 grid results, the activity levels steadily dropped. What use would MSE be if the activity levels were so low as to be undetectable?

The extension of the MMA model to larger senses quickly shows that there is something missing from the MMA model. With each increase in sense size, the signal to noise ratio drops substantially. Tuning of the sigmoidal curve is not enough to fix the problem. Both the experiments with larger senses and the initial 2x2 sense extension point to a need for inhibition within the MMA framework.

5.3.4 MMA Model with Inhibition

In my third extension of the MMA model, I added a form of inhibition into the SOM such that the winner and its three nearest neighbors silenced the activity of all other neurons. (The distance of three was chosen so as to not interfere greatly with the neighborhood update function.) The same parameters used in the last extension (2x2, 3x3 and 4x4 grids with tonic

bias set according to MMA procedure). As you can see in Figure 44, there are areas of blue/red that represent MSE surrounded by large areas of green representing MSD. There was one special consideration in the calculation of MSE: the case when the unisensory responses were 0 (inhibited) and the multisensory response was non-0 (not inhibited). MSE is infinite in this case, but I set these values to have the value of 1000. As the graphs show, there is no more MSE generated by noise level activity. In fact, noise level activity is turned off by the active population. This is how the networks in our brain work; inhibition keeps the activity levels from spiraling out of control and causing seizures. This was a first step to show it is possible to incorporate inhibition into the MMA work to counteract the effects of decreasing signal to noise ratio. A more realistic setup would be to use Ursino et al.'s Mexican hat in SOM's activation functions.

5.3.5 MMA Model with Overlapping Receptive Fields

The final extension I added to MMA's model was the incorporation of overlapping receptive fields. Instead of the world stimulating each sensory neuron individually, it stimulated overlapping groups of them. The protocol I used for the results I will show was: a location stimulates the neuron in the corresponding location with the driven probability and the nearest neighbor neurons (including the diagonals) were stimulated with a weakly driven probability. No matter what I did with the setup, MSE would not form in the same way that it formed for MMA.

In Figure 45, I used the same parameters for the 3x3 senses except that one had a weakly driven probability of 0.3 whereas the other did not have weakly driven activity. The latter forms a SOM and shows MSE in the same manner as MMA, but the addition of weakly driven activity changes the properties of the SOM and MSE in the former case. Weakly driven activity causes the SOM to try to map both the relationship between inputs of a particular modality and the relationship between inputs of different modalities. The result is a spatial map that is somewhere in-between the two, but that can no longer work as a multisensory integrator. I tried different combinations of parameters (changing accuracy, size of the senses and the strength of the weakly driven activity), but the results were the same. This was the biggest limitation of the SOM based model of multisensory integration.

If we take a closer look at the first extension I performed, MSE occurred within the same sense if both unisensory stimuli were applied to one sense. This gets at the same problem as with weakly driven activity. A SOM is an organizer of information, not an integrator of information. It is only a special case of using a SOM under controlled conditions that turns it into an integrator. Recall the function of a SOM is to sort out inputs into different regions of the map based on vector similarities. Unfortunately, the differences between the vectors that represent different information from one sense (like two visual targets in your receptive field) are the same as the differences between the vectors that represent different sensory information (like a visual target and an auditory target). In other words, the SOM cannot distinguish the information within a sense from the information across senses.

To illustrate, let's return to the 2x2 sense experiment. Figure 46 shows a simplified illustration of the weight distribution from that experiment. Each corner of the grid responds to a particular case of unisensory inputs (the upper left responds to the case when the first sense responds to a vertical stimulus; the upper right responds to the case when the second sense responds to a horizontal stimulus; etc.). The purple lines indicate the gradients between unisensory areas. These are a result of the neighborhood function of the SOM. If a point

responds to a particular stimuli, nearby areas respond to similar stimuli. Figure 47 shows the overlap between the two unisensory neighborhoods that results in MSE. Because the SOM does not discriminate between sensory information, this same overlap occurs between the information within a sense- the upper left and the lower right (See Figure 48).

The SOM algorithm is an organizer not an integrator. The integration occurs as a result of where particular pieces of information are placed. Unfortunately, the SOM does not discriminate between sensory information, so information within a sense will generate MSE just as information between senses.

6. Theoretical Model

6.1 Problems with SOMs

All of the models of multisensory integration have their weakness. The problem has been that there is no foundation upon which these models rested. They each pick the evidence they want to fit and come up with a way to fit it. The MMA model was the first to step back and try to get at the fundamental unit. Their model is not without problems, but it can be used as a starting point to build back up to the complex models like the Ursino and AP models.

What MMA does well is explain a way to generate MSE using a simple, biologically-inspired mechanism. My research shows that SOMs are not the end all answer to multisensory integration. Any movement to a more complex sensory apparatus than point sources results in the breakdown of the model. The main problems can be divided into problems with the sigmoidal function, problems with signal to noise and limitations of 2D SOMs.

The sigmoid problem is that the values for the sigmoid parameters have not been worked out biologically. Instead, the MMA model set their tonic bias of the sigmoid to what would produce MSE where they wanted it in a system. Their results are more a proof-of-concept than evidence that this is necessarily how the brain works. The same critique applies to the Ursino and AP models. Without a biological basis for the values, a host of values can be deemed reasonable although they can produce quite different results (as show in my experiments). Essentially, the areas that show MSE can be controlled by the sigmoid parameters. Unfortunately, there is no solution to this problem from the computational side. More detailed data would need to be obtained from in vivo electrical recordings to better understand the threshold function of neurons involved in multisensory integration.

The signal to noise problem is that as the size of the senses is increased the noise levels start to creep into the areas of MSE. This effect can be nullified by altering the sigmoid, but this dampens the unisensory responsiveness of the network. A simple solution would be to manually force the noise parameters in the model to lower values. The biological mechanism that would correspond to that is inhibition. As my results showed, even a rudimentary inhibition function has large gains in restricting MSE to the multisensory areas through managing the signal to noise ratio. Inhibition is essentially to brain function, so it is a logical and necessary extension of the MMA model.

The biggest problem is that a SOM cannot map more than one relationship in the data effectively and still show MSE in the right places. The MMA model limited itself to exclude the effects of receptive fields (by having each modality have the same receptive field). My experiments show that SOMs alone are not enough to map both within-sense information and between-sense information. The receptive fields of the sense have to be limited such that there is no relationship among the inputs of a single sense. Even when that is done, something is still

missing from the model because of the MSE within a single sense in my results. Inhibition is the key here as well. My inhibition function works by selecting the winning population and silencing all other neurons. A more realistic inhibition function would have a Mexican distribution of lateral weights (like the Ursino model). Every neuron would excite and inhibit the other neurons in the network. This would allow for dynamic competition wherein two unisensory stimuli inputs to the same modality would silence each other. If both populations are silenced, no MSE can occur.

6.2 New Model

My experiments showed the limits of the MMA model in isolation of the other models. The power of modeling is the ability for parts of models can be swapped out and combined. The MMA SOM algorithm alone does not tell the whole story of MSE, but they did identify a fundamental computational unit that could be used by the brain if certain precautions are taken. First, the within sense relations must be coded separately from the between sense relations. Second, inhibition must be involved to keep the signal to noise ratio in check and to explain MSD. The MMA model can be envisioned as the central part of a training model version of the Ursino Model.

Following the Ursino model setup, there would be three layers: two unisensory layers and one multisensory layer. Each layer would be broken down into areas that respond to a narrow slice of the world. These areas would be arranged topographically. In the unisensory layer, these areas would be single neurons, but in the multisensory layer, these areas would be a group of neurons trained by a SOM. The corresponding neurons from each unisensory layer (those that have the same receptive field) will project to the same multisensory area. Instead of hardwiring these connections like the Ursino work, the MMA algorithm would be applied to train the weights. This will produce the robust MMA MSE results in the context of a more realistic hierarchy.

The Mexican hat distribution of weights between neurons would be applied to each area in the layers. For the unisensory layers, this means that each neuron will have the Mexican hat distribution. An active neuron will excite nearby neurons, while silencing other neurons. For the multisensory layer, the same concept will be applied except that the areas will be the base unit. An active area will excite nearby areas, while silencing other areas. This will incorporate the important Ursino findings on suppression.

6.3 Future Steps

In the prior section, I state that unisensory areas correspond to single neurons. My work shows that that may not be necessary; unisensory areas could correspond to clusters of neurons if precautions are taken. Incorporation of inhibition within these areas solves both the problem of decreasing signal to noise ratio with bigger senses and the problem of MSE within a single modality.

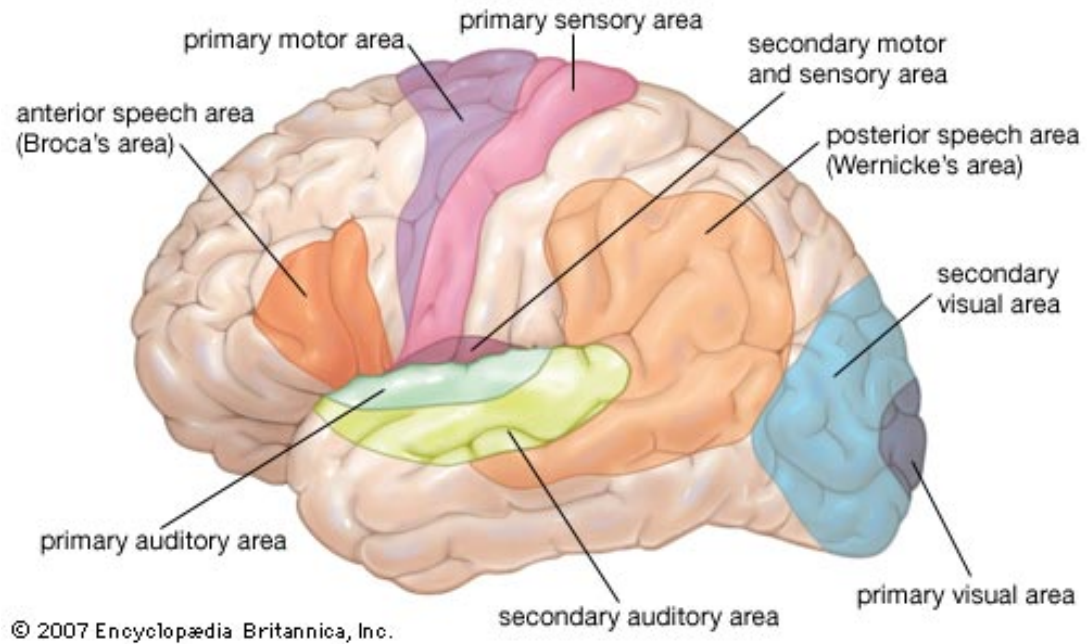
As a first step, the topographical organization of each layer could be hard coded like the Ursino model did. In reality, there should be two SOM algorithms operating at two different levels. The first is the MMA procedure to generate MSE when individual areas are combined. The second would organize the areas within a layer. Unfortunately this second SOM is more complicated. Organizing each layer independently is not a problem, but lining up each layer so that the right areas project together (i.e. the receptive fields line up) is a more complicated problem. I believe a feedback mechanism would be necessary to solve this problem.

Unfortunately, this extension is beyond the scope of this thesis, but would be a future modification that could be made to the model. If it were added, the explanatory power of the model would increase. Not only would the model account for MSE and MSD, but it would also give a plausible mechanism by which the system develops.

Another modification that would help fit the biology would be the incorporation of ascending and descending projections like in the AP model. Once the validity of the original model is established, this extension could be build upon the foundations.

6.4 Conclusions

The literature from neuroscience and psychology can only move our understanding of multisensory integration only so far because of limitations in technology. Modeling allows us to test different theories of how the mechanisms of integration work. The MMA model constitutes a proposal for a most basic unit of multisensory integration. This fundamental unit can then be used to construct more complete models of multisensory integration. My theoretical model is based on the superior colliculus literature, but is not specific to that setting. It is the first step in allowing us to model and test the hypotheses surrounding the cortical literature.



**Figure 1: A Diagram of Sensory Brain Areas
(Reproduced form Britannica Online)**



Figure 2: Rubber Hand Illusion

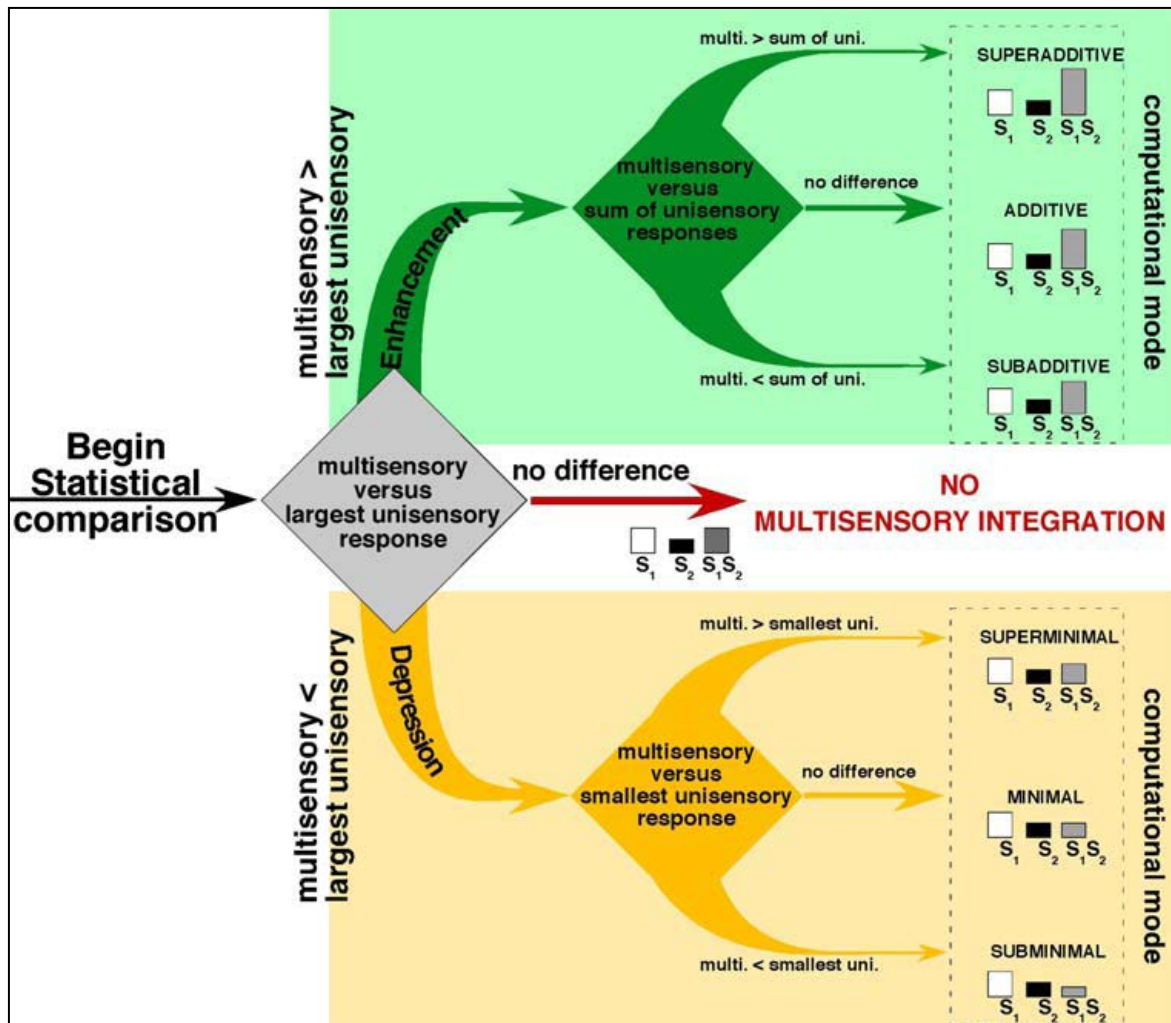


Figure 3: Types of Multisensory Integration from a Neuron's Perspective

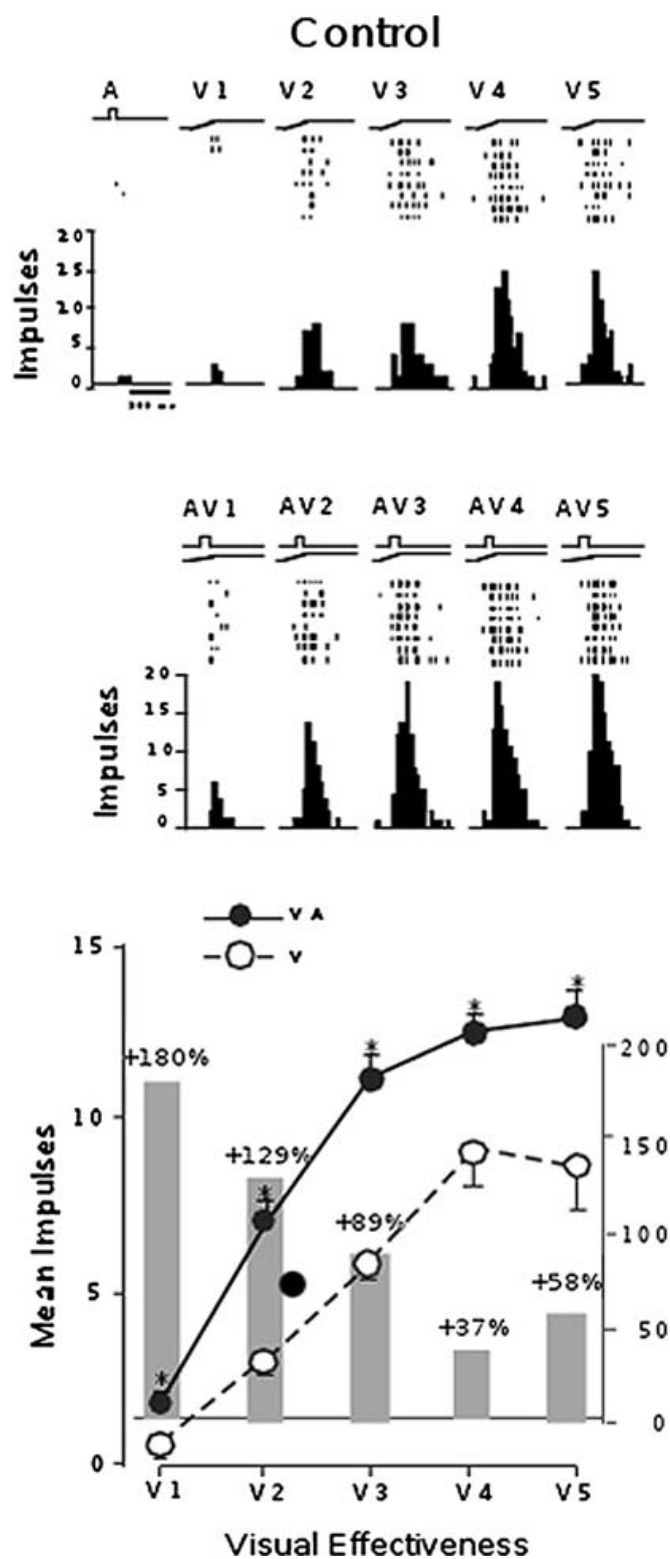


Figure 4: Inverse Effectiveness in a Neuron

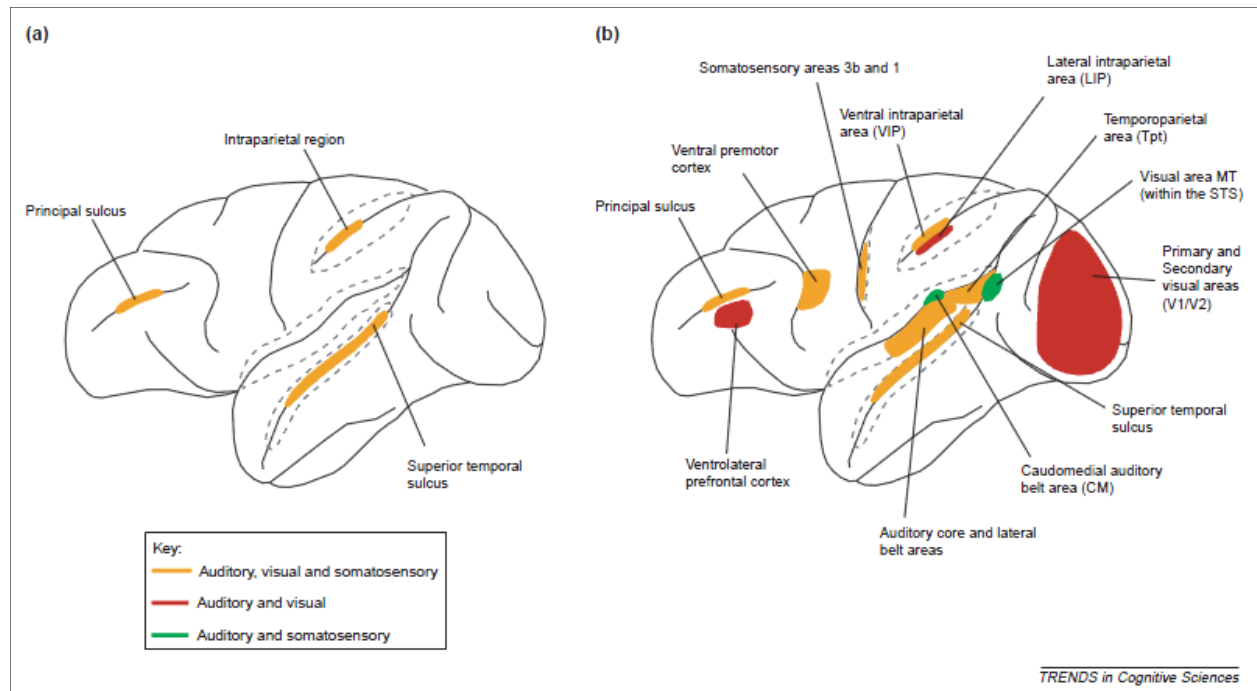


Figure 5: Areas of the Cortex Receiving Multisensory Information

Model	Results	Essential Components	Weaknesses
Rowland, Stanford and Stein	<ul style="list-style-type: none"> •MSE •Superadditivity •Inverse effectiveness •Cortical deactivation 	<ul style="list-style-type: none"> •Modulatory role of descending connections •Inhibition 	<ul style="list-style-type: none"> •Assumptions about inputs •Modeled at cellular level
Ursino, Cuppini, Magosso, Serino and Pellegrino	<ul style="list-style-type: none"> •MSE •Superadditivity •Inverse effectiveness •Multisensory response time •Cross-modal suppression •Within-modal suppression 	<ul style="list-style-type: none"> •Sigmoidal activation •Mexican hat •Hierarchy 	<ul style="list-style-type: none"> •Everything is hard-wired •Only single set of parameters tested •No descending/modulatory connections
Anastasio and Patton	<ul style="list-style-type: none"> •MSE •Superadditivity •Inverse Effectiveness •Cortical deactivation •Developmental timeframe 	<ul style="list-style-type: none"> •Sigmoidal activation •Primary connections extract information •Modulatory connections integrate 	<ul style="list-style-type: none"> •No inhibition •No MSD •Assumptions about inputs •Simplified modalities
Martin, Meredith and Ahmad	<ul style="list-style-type: none"> •MSE •Superadditivity •Inverse Effectiveness 	<ul style="list-style-type: none"> •Sigmoidal activation •SOM training 	<ul style="list-style-type: none"> •No inhibition •No MSD •Simplified modalities •No descending/modulatory connections

Figure 6: Table of Computational Models

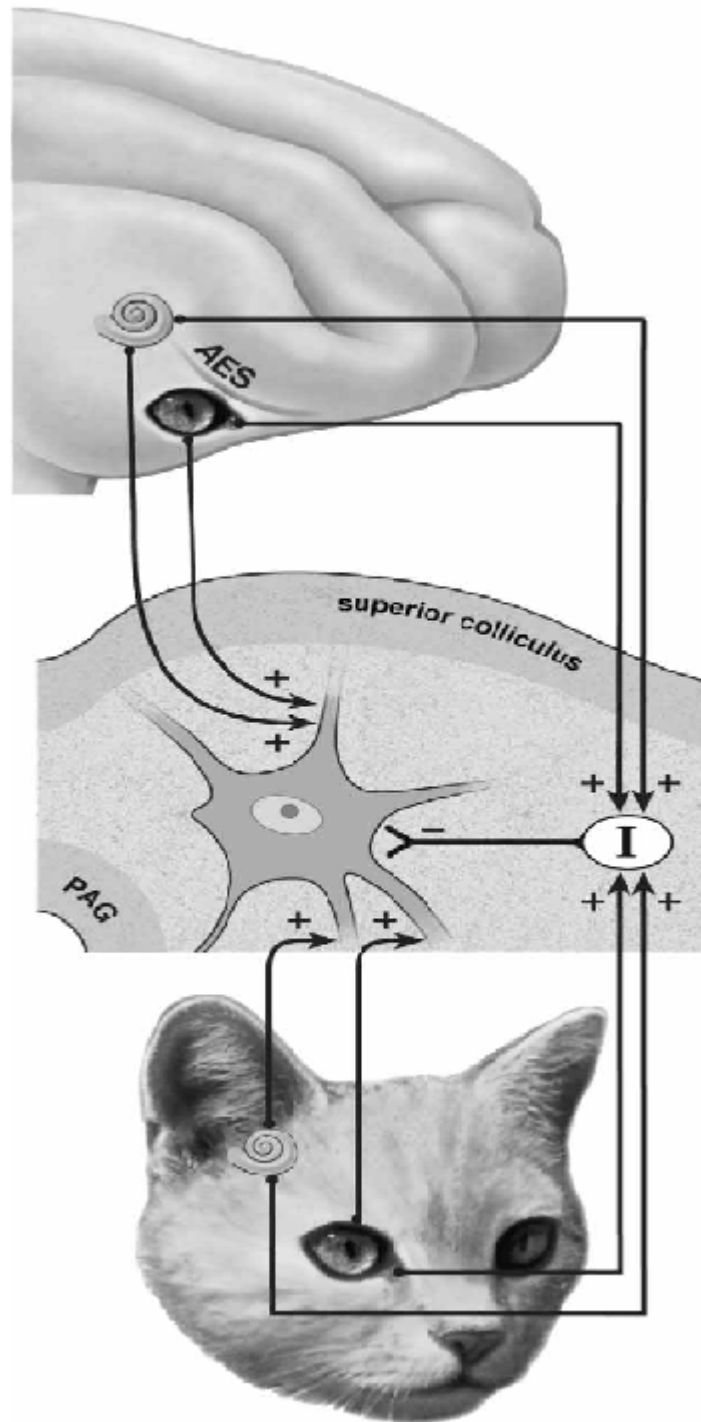


Figure 7: Rowland and Colleagues' Schematic for their SC Model

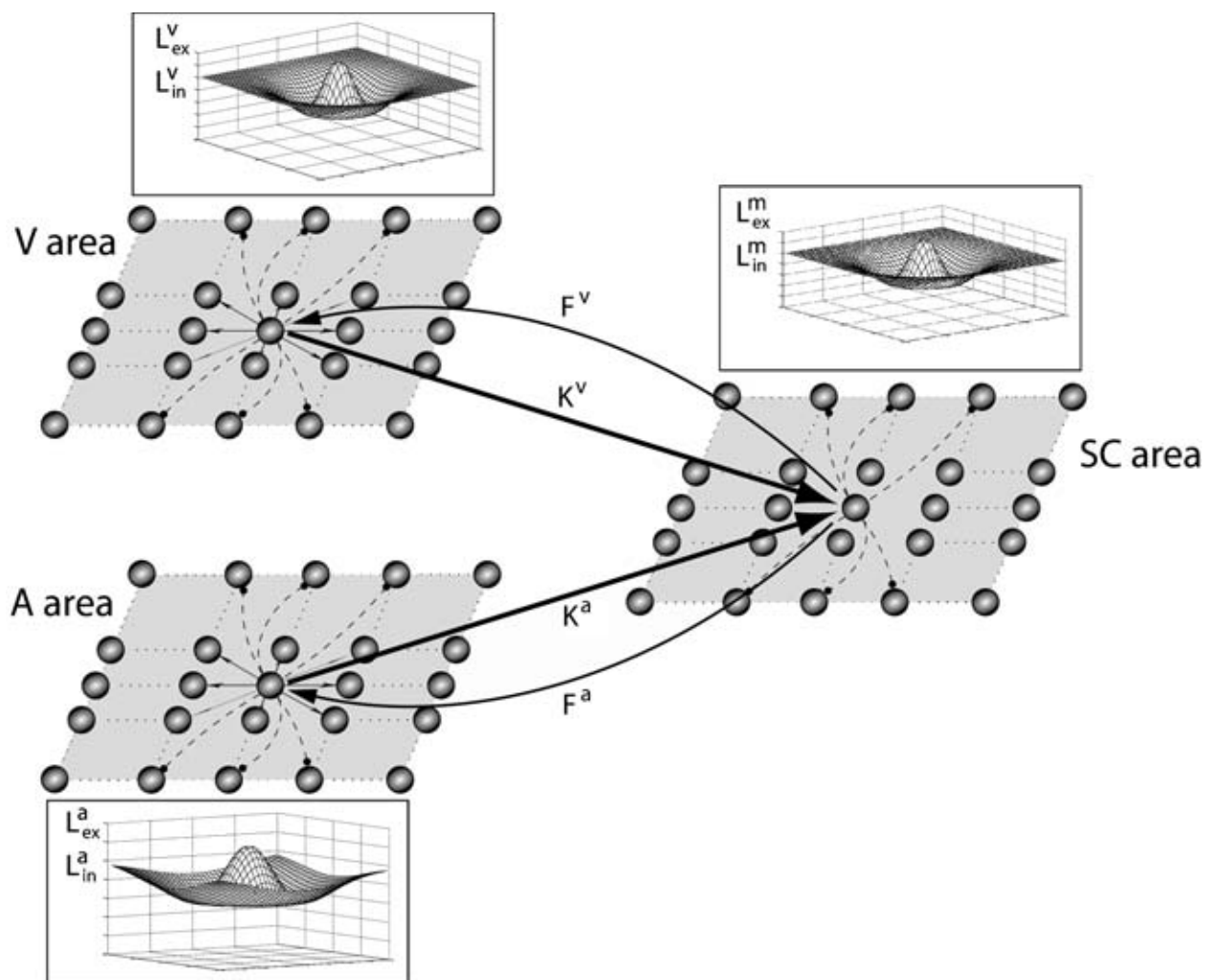


Figure 8: Ursino et al.'s SC Model hierarchy

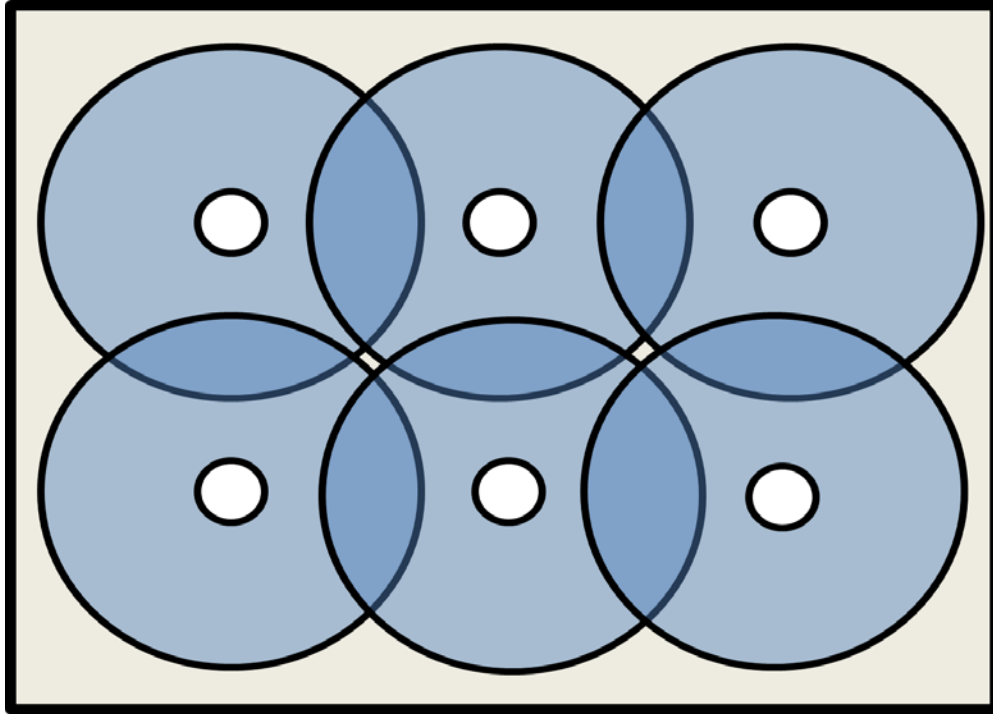


Figure 9: Overlapping Neural Receptive Fields

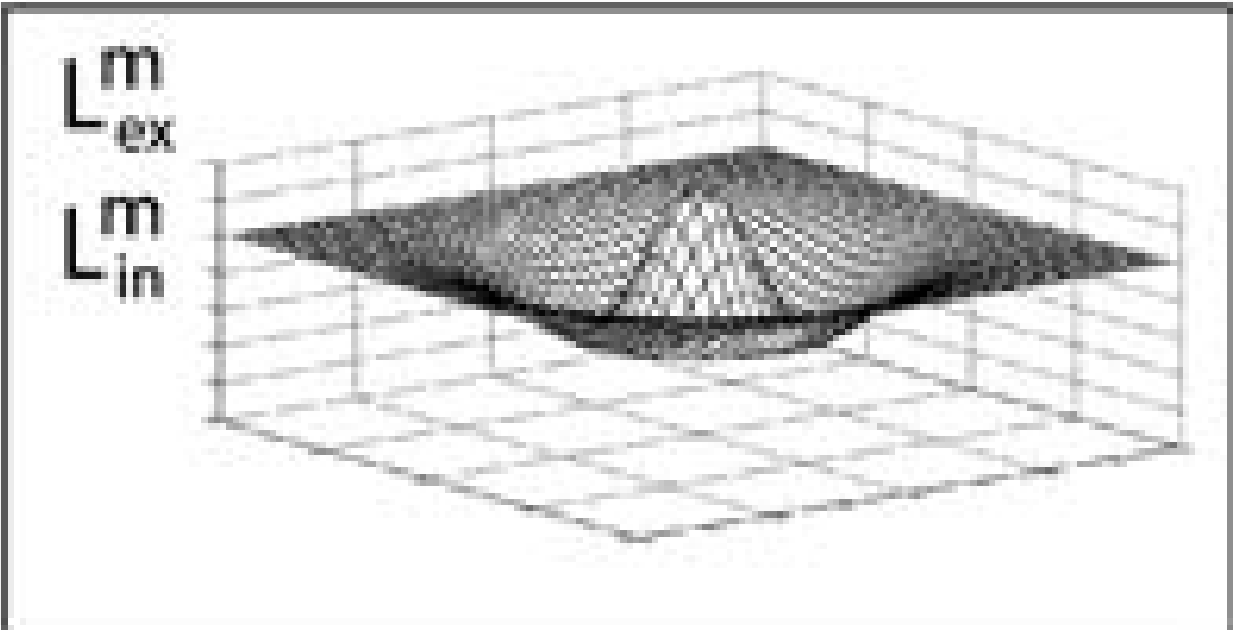


Figure 10: Mexican Hat Distribution of Weights

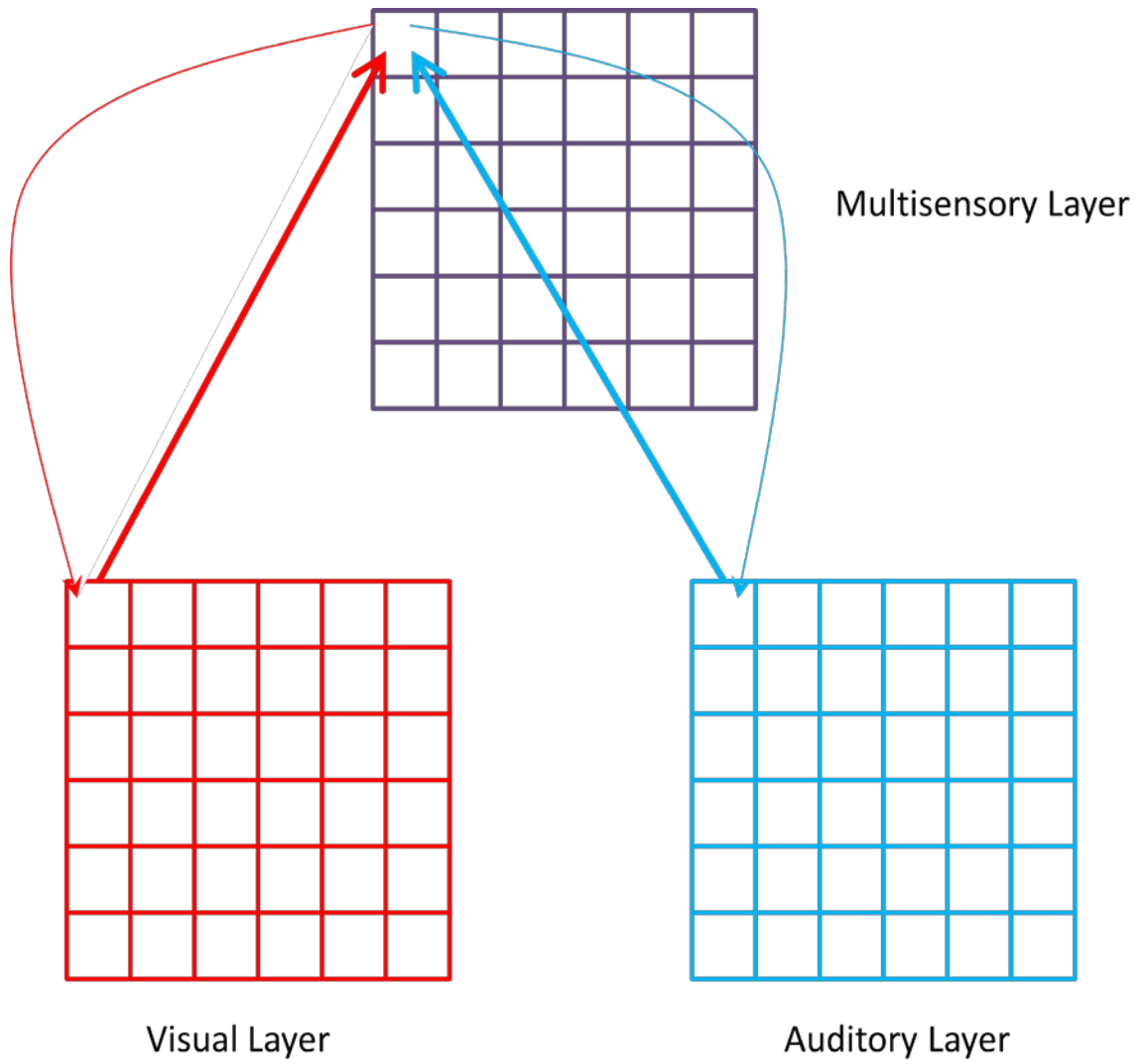


Figure 11: Ursino et al.'s Connection Scheme

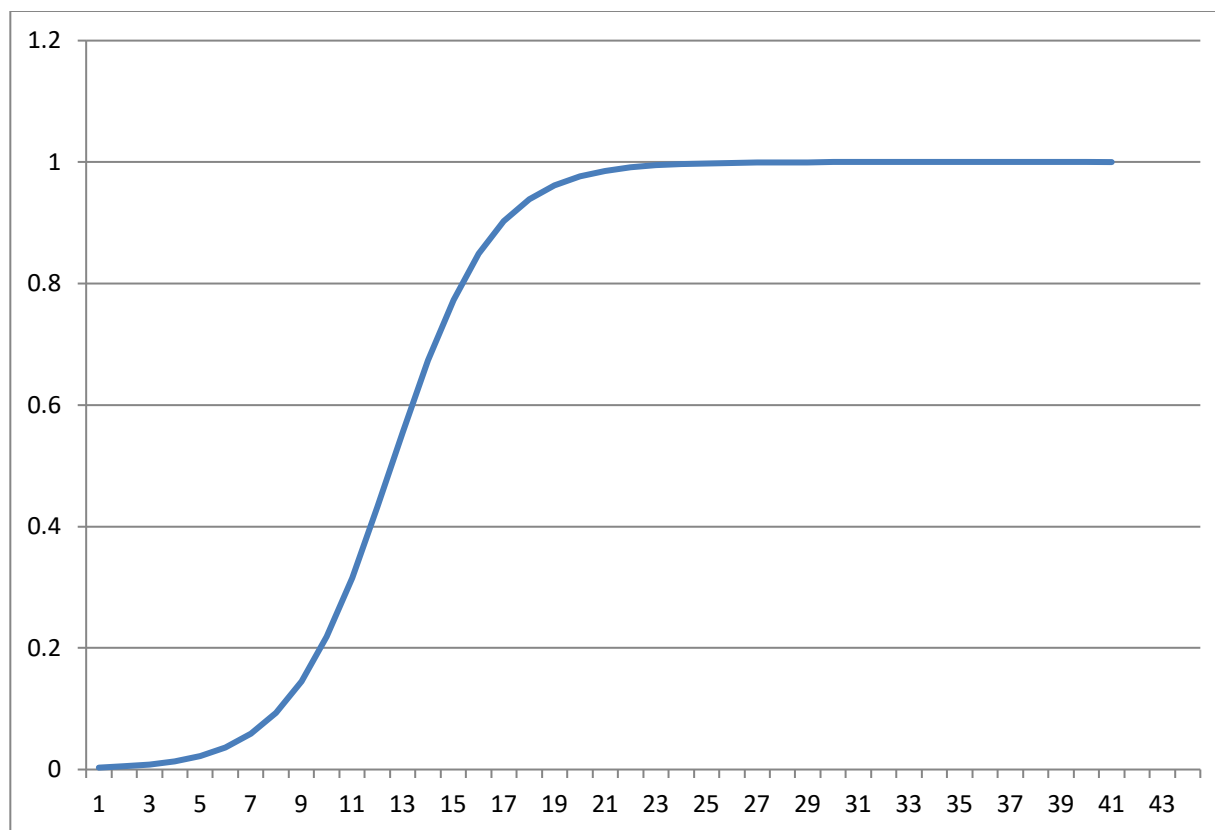


Figure 12: Example Sigmoidal Curve

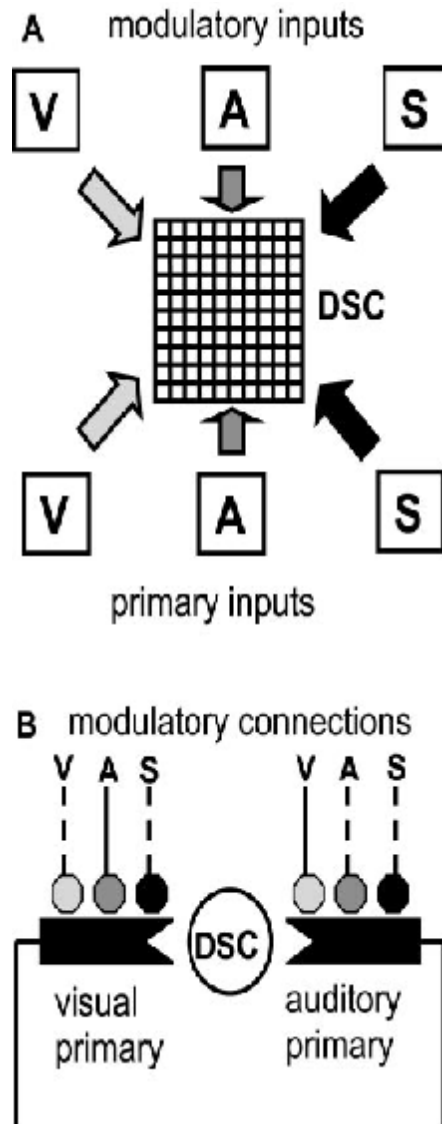


Figure 13: Anastasio and Patton's Schematic for their SC Model

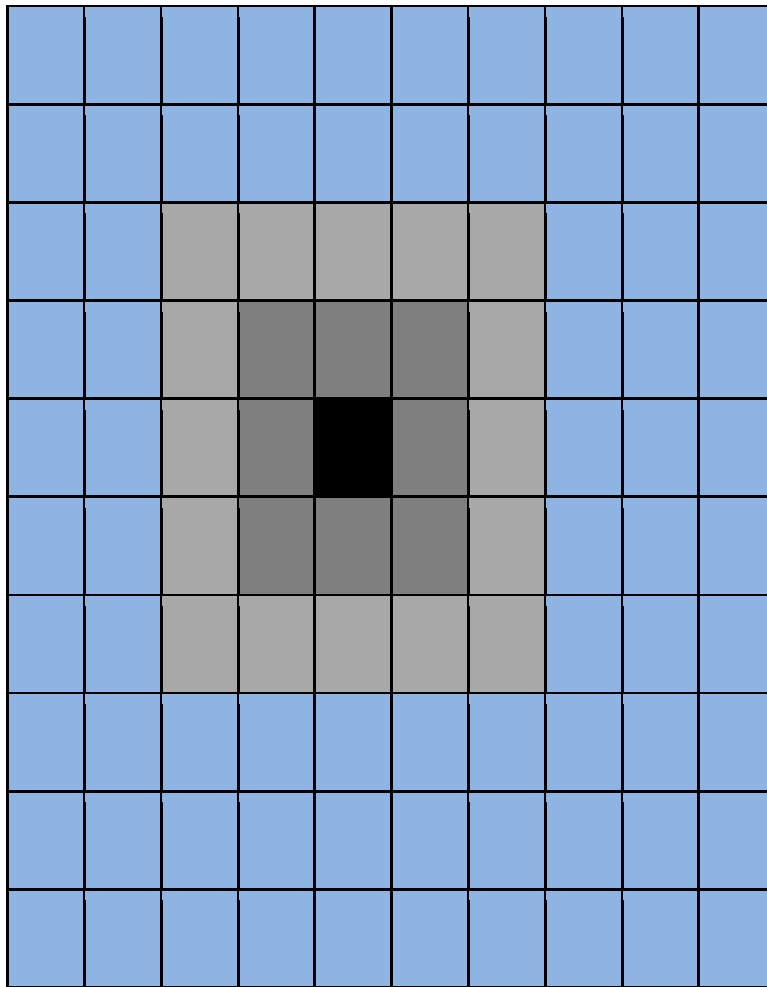


Figure 14: Illustration of Winner and Neighborhood Selection

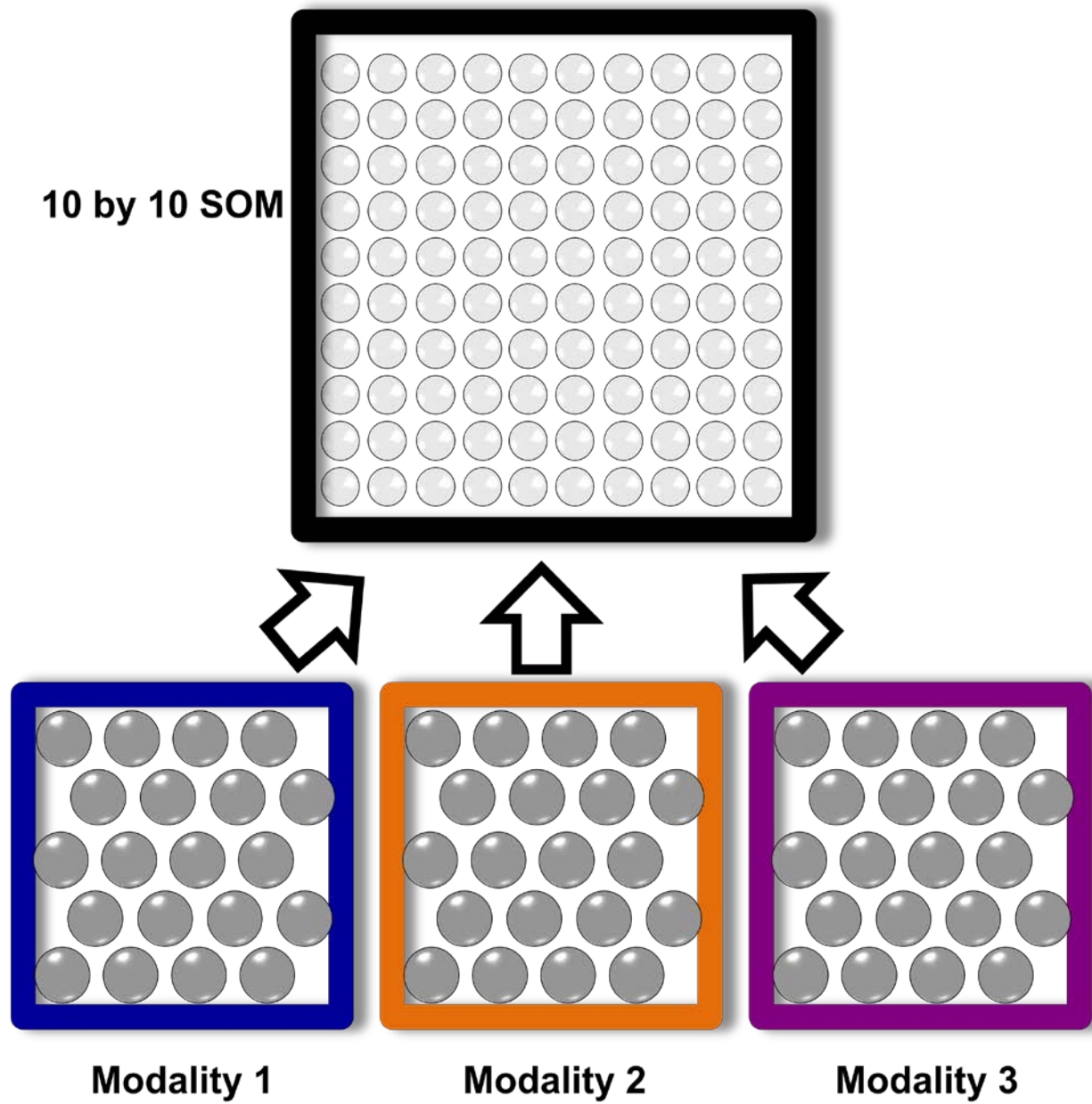


Figure 15: Illustration of MMA's model

Unimodal	001	010	100
Bimodal	011	101	110
Trimodal	111	000	

Figure 16: Table of Modality Bitstrings

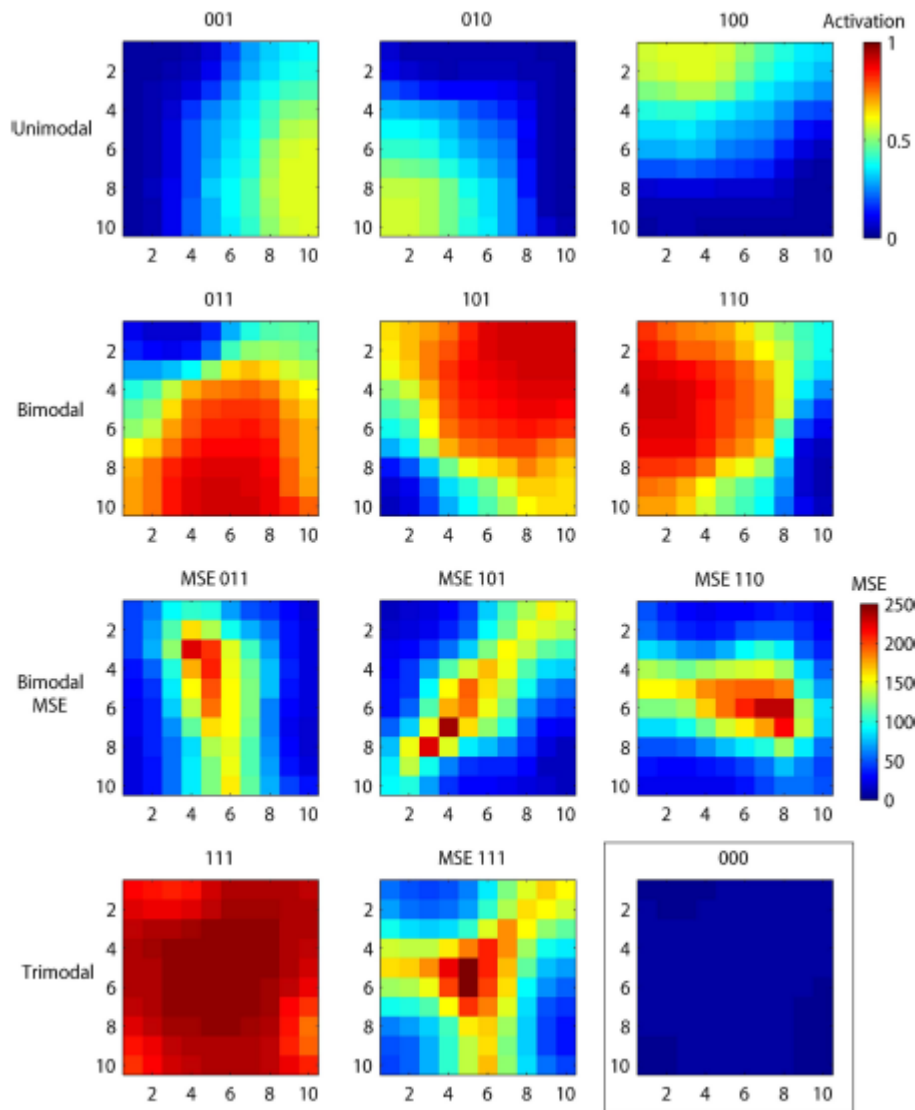


Figure 17: The MMA Model's Results

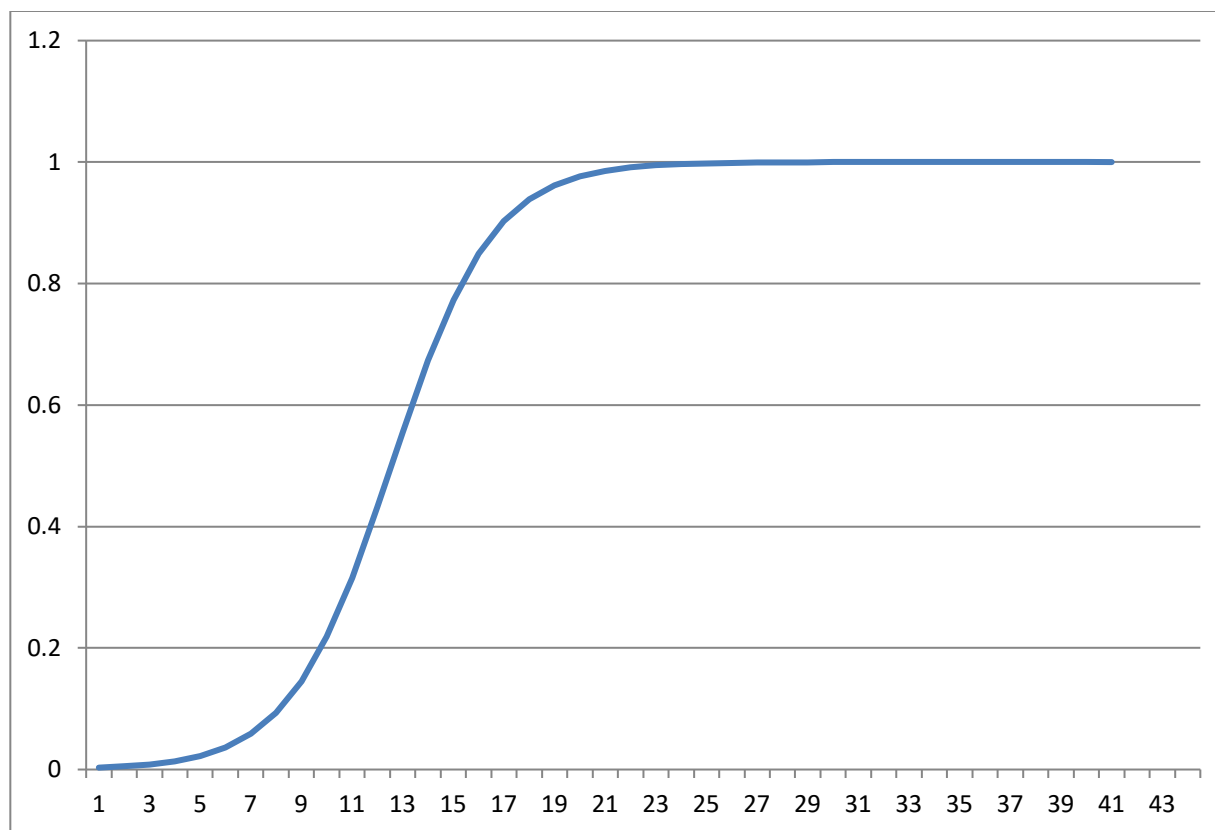


Figure 18: MMA's Sigmoidal Curve

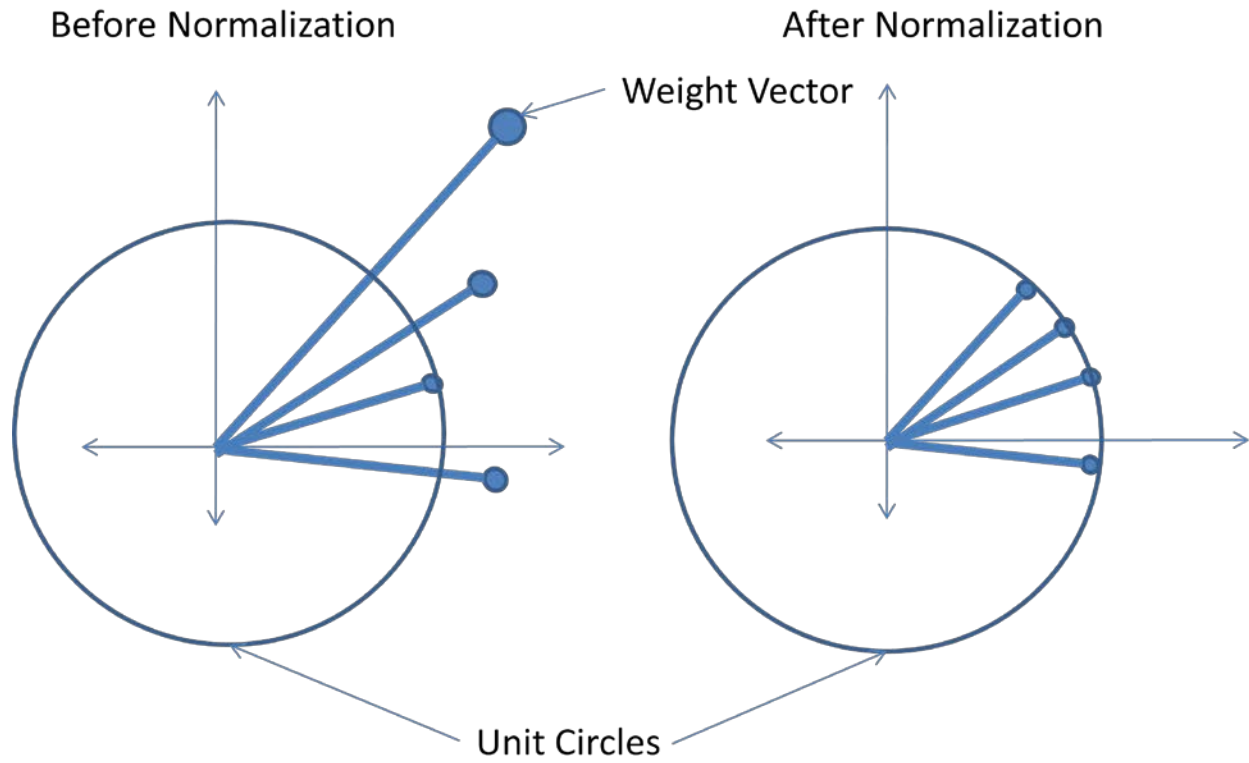


Figure 19: Illustrations of the Effects of Normalization

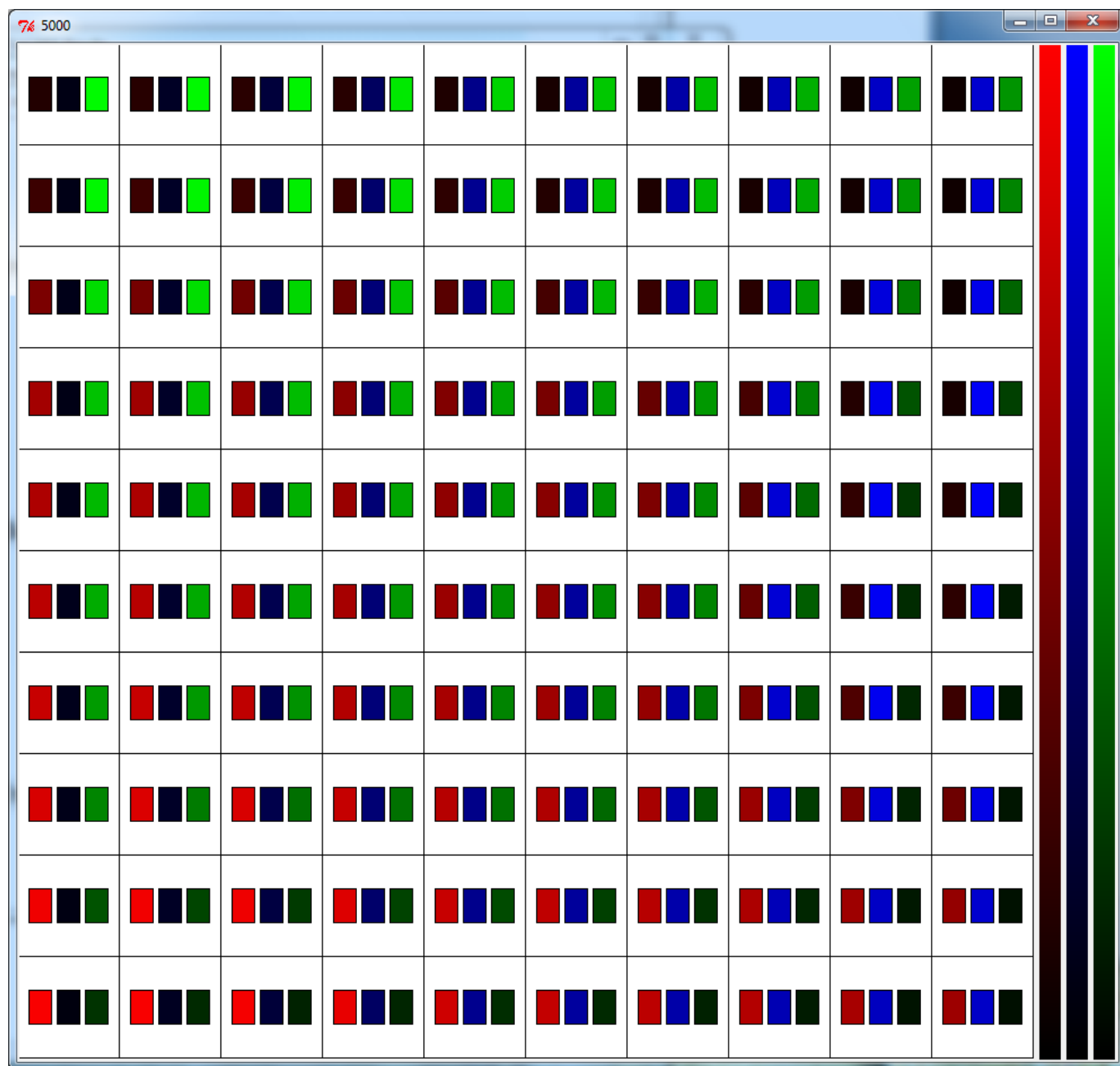


Figure 20: Weight Distribution in Replication of MMA

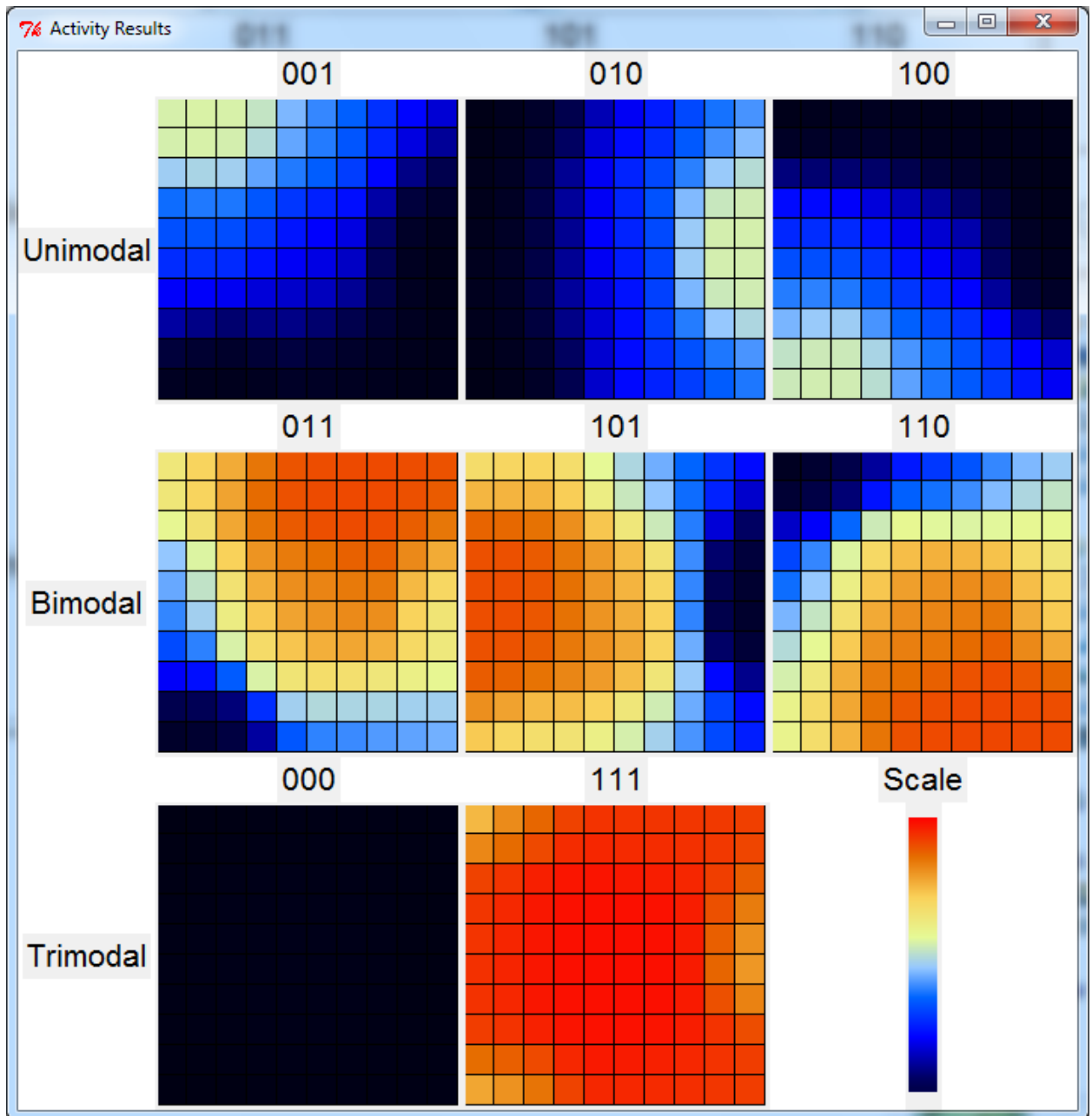


Figure 21: Activity Averages in Replication of MMA

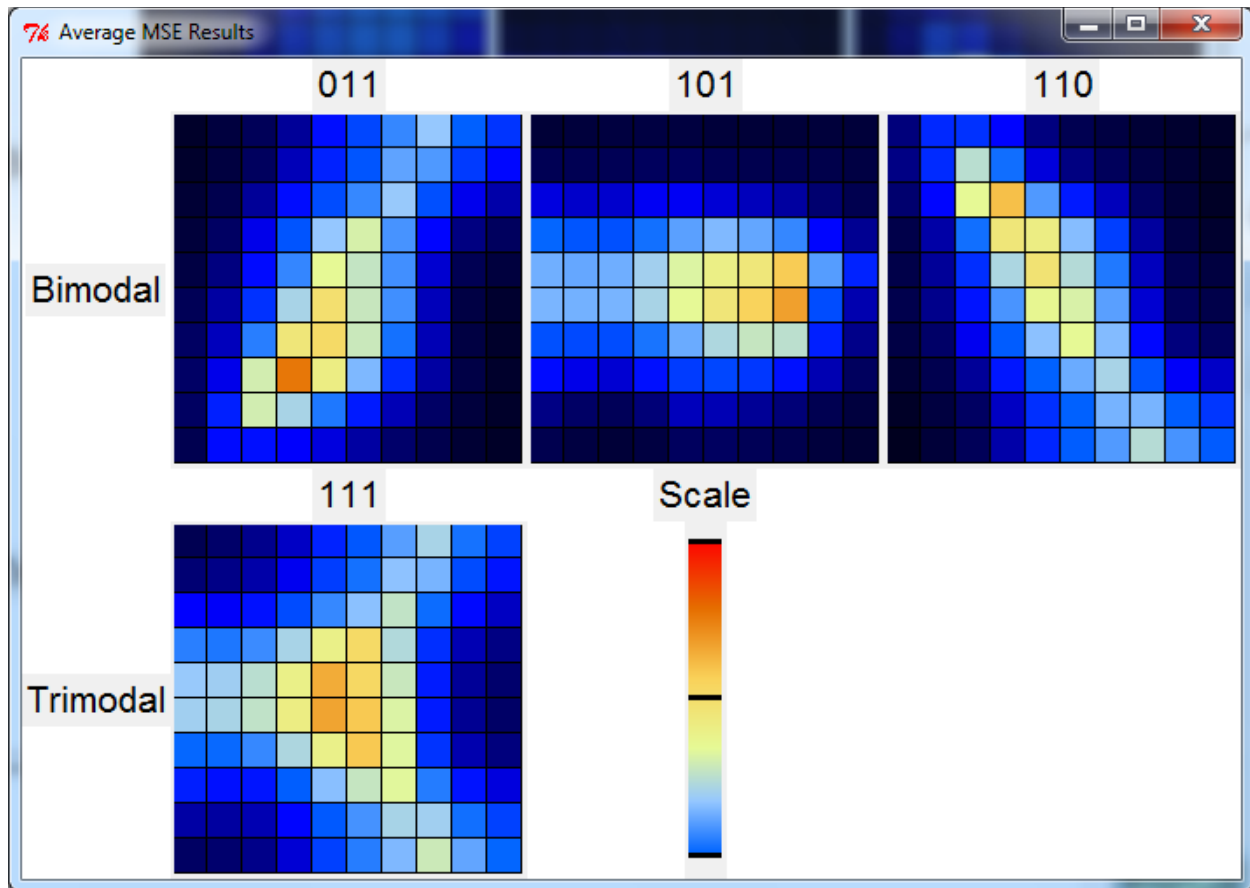


Figure 22: MMA's Average MSE Metric

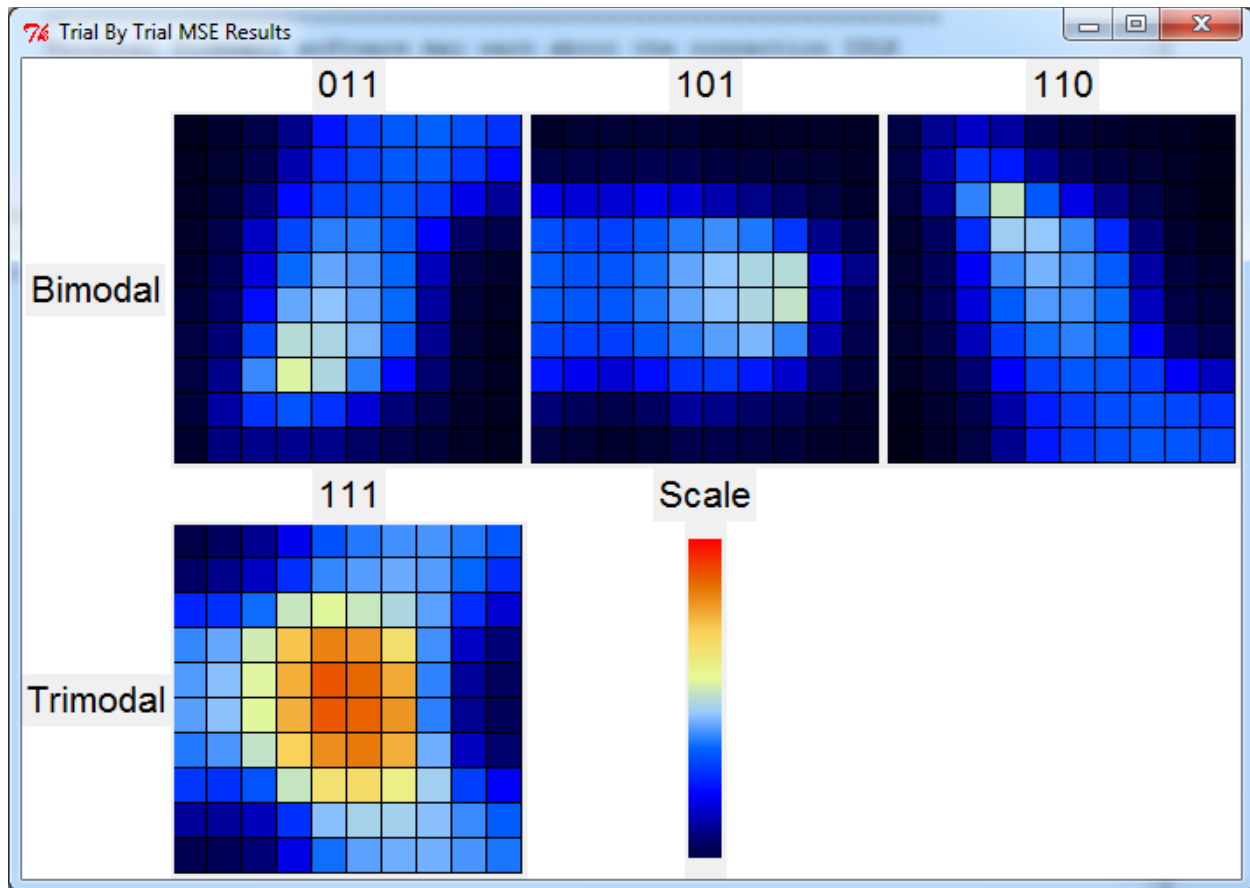


Figure 23: My Average MSE Metric

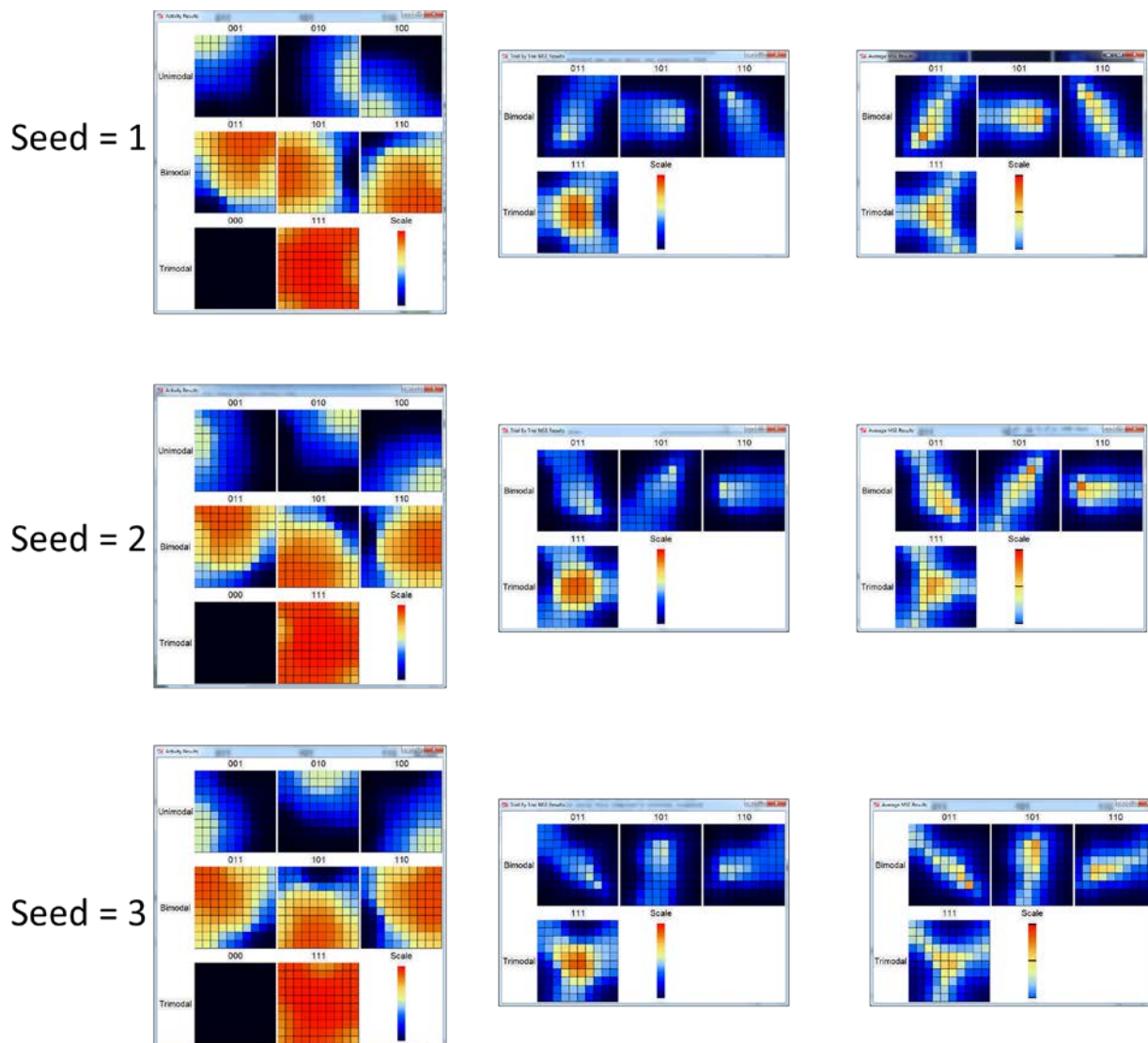


Figure 24: Same Trend Over Different Random Seeds

	Unisensory Activity	Multisensory Activity	MSE
	0.1	1.0	900%
	0.3	0.5	66.7%
	0.6	0.75	25%
Average	0.33	0.75	330.7%

$$\% \text{ MSE} = (\text{MS} - \text{US}) / \text{US}$$

Average MSE according to the MMA paper's protocol:

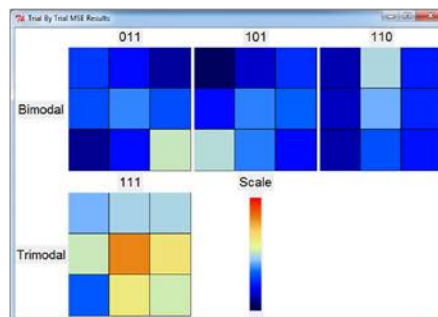
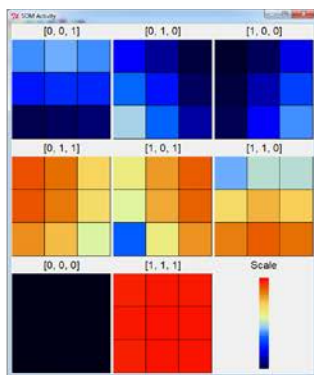
$$(0.75 - 0.33) / 0.33 * 100 = \mathbf{127.3\% \text{ MSE}}$$

Average MSE as calculated by averaging the individual MSE values:

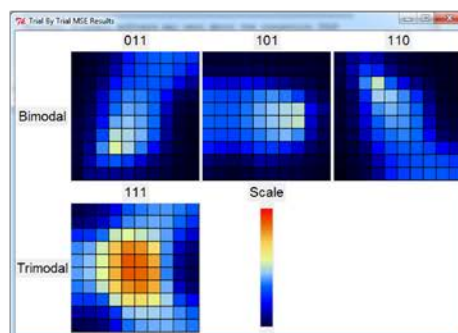
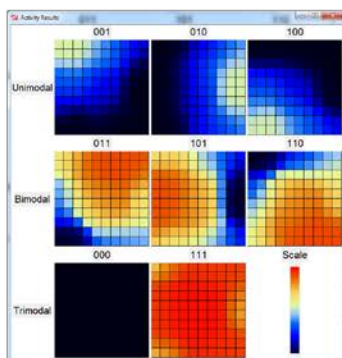
$$(900\% + 66.7\% + 25\%) / 3 = \mathbf{330.7\% \text{ MSE}}$$

Figure 25: Comparison of MSE Metrics

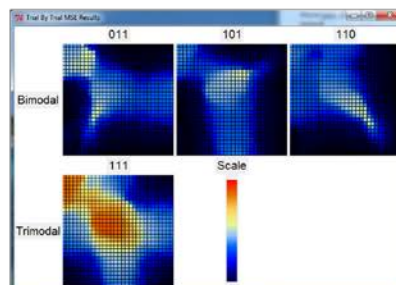
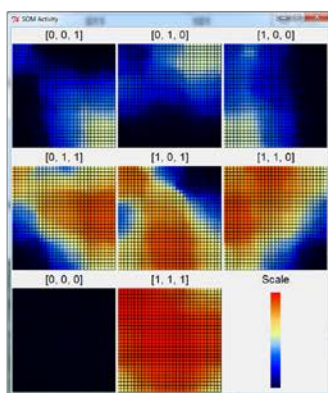
3 x 3



10 x 10



30 x 30



100 x 100

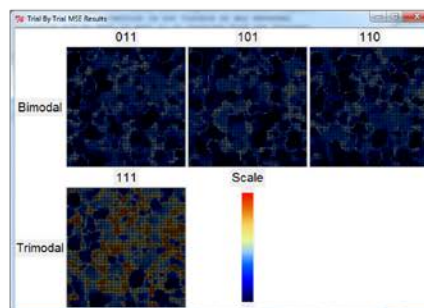
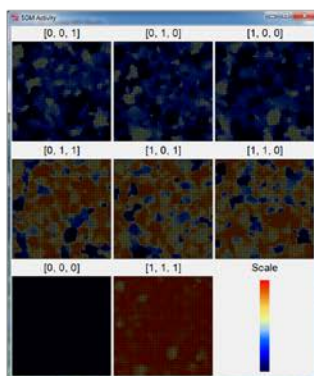
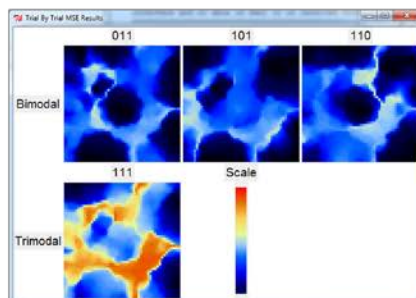
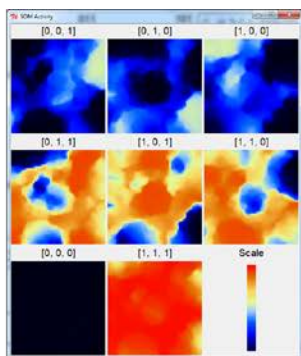
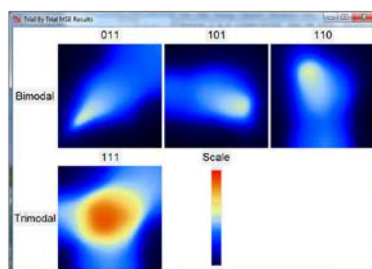
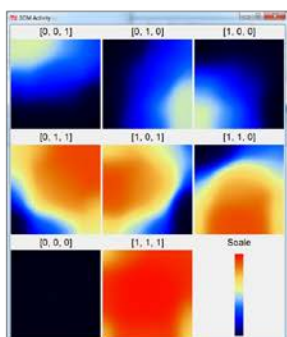


Figure 26: Size Trials

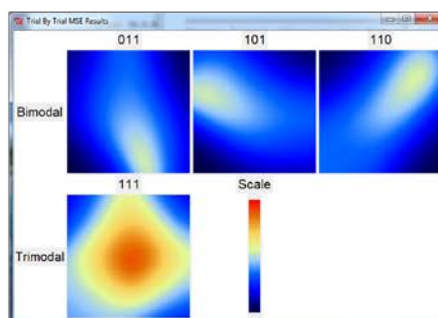
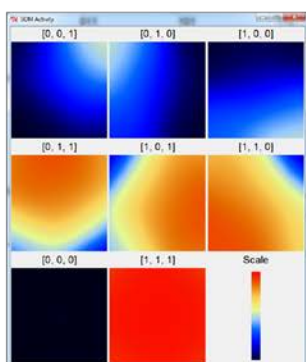
1



5



15



30

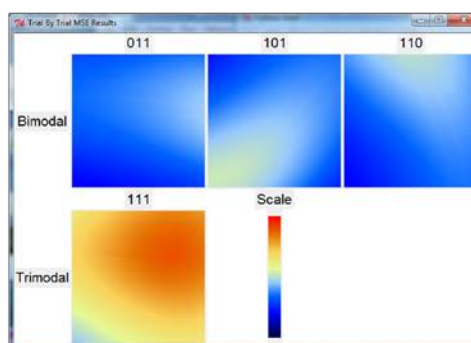
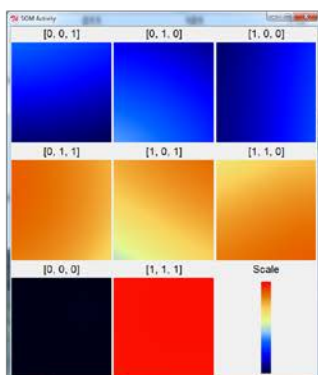


Figure 27: Width Trials

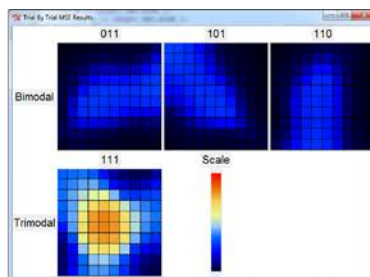
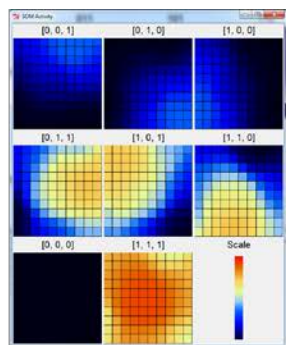
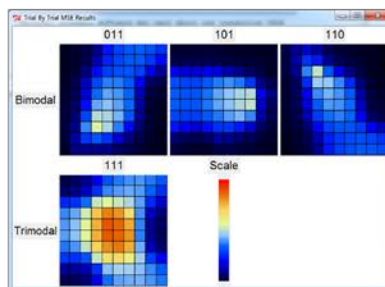
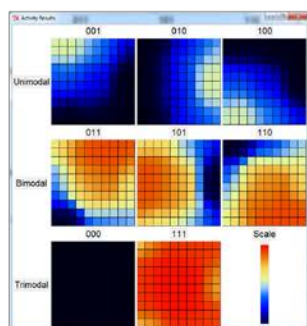
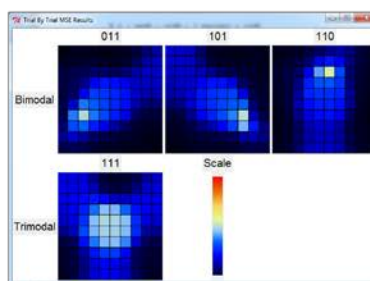
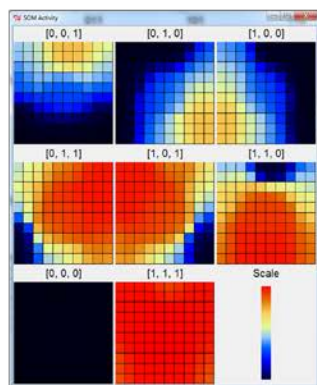
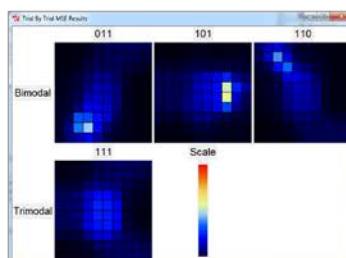
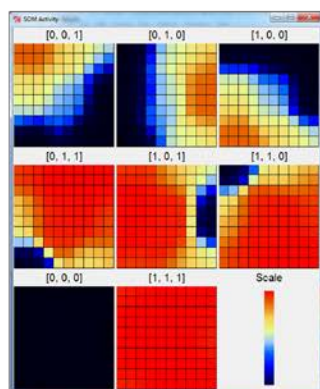
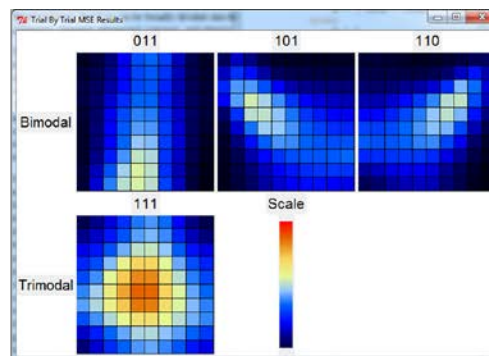
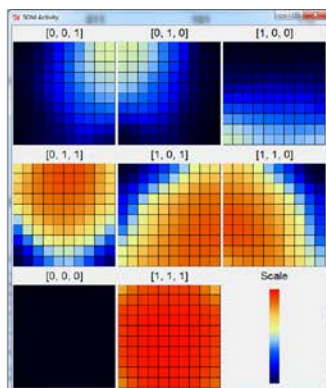
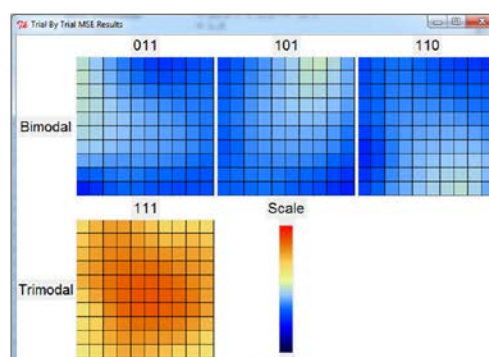
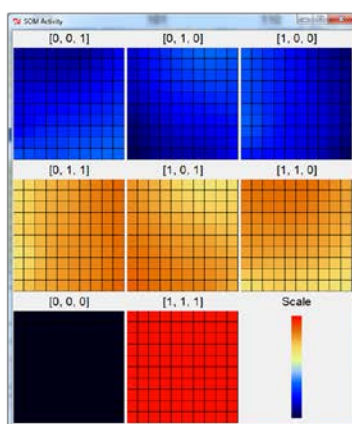
$P_d = .5$  $P_d = .6$  $P_d = .7$  $P_d = .8$ 

Figure 28: Probability Driven Trials

Only 000



Only 111



Only Unimodal

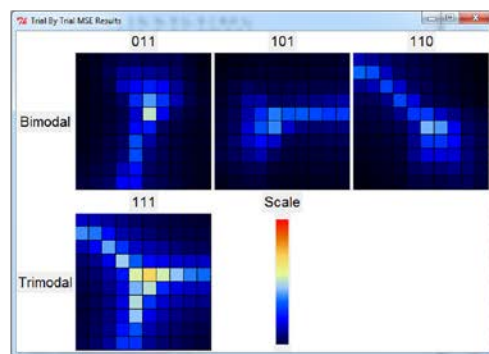
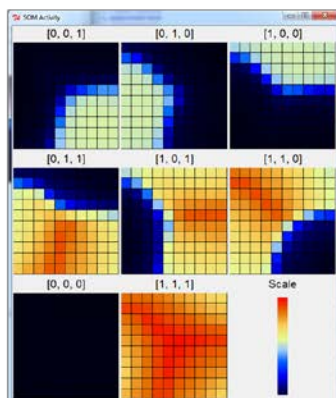
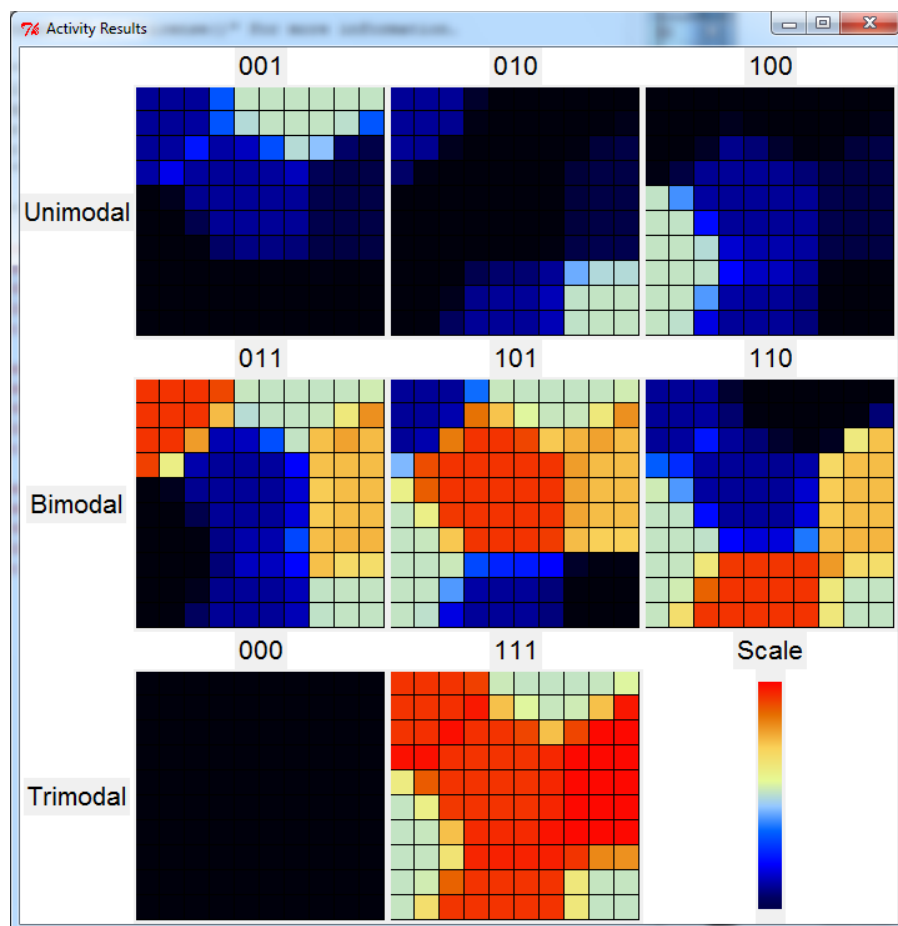


Figure 29: State Probability Trials

**Figure 30: No Noise Trial**

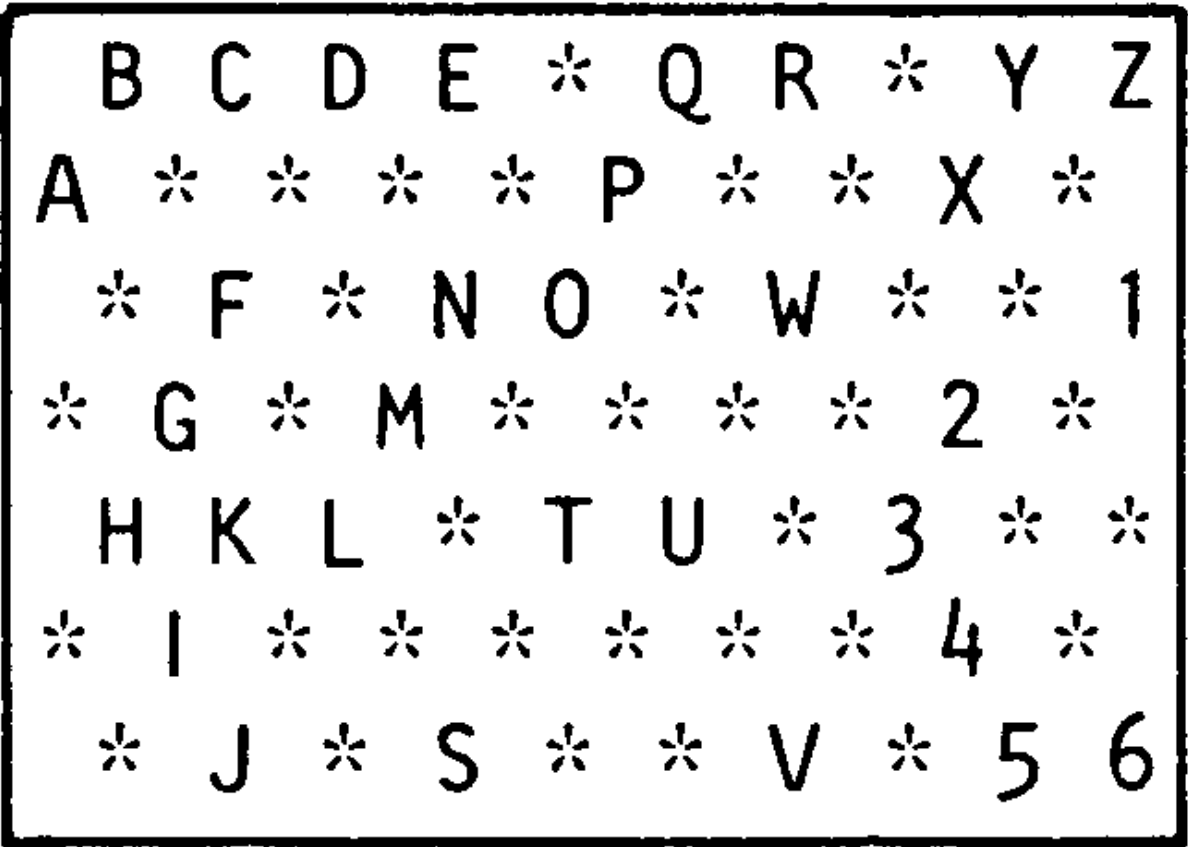
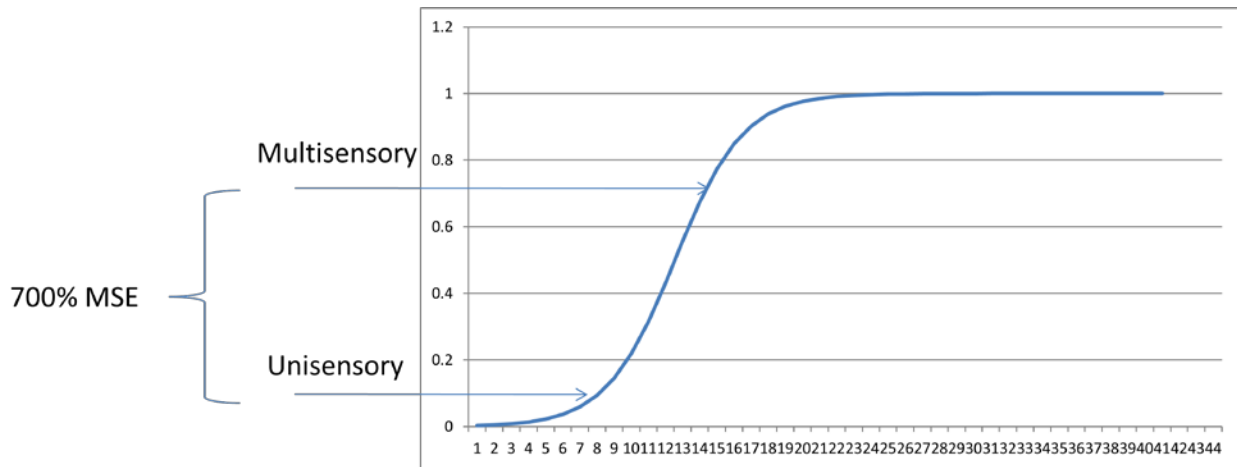
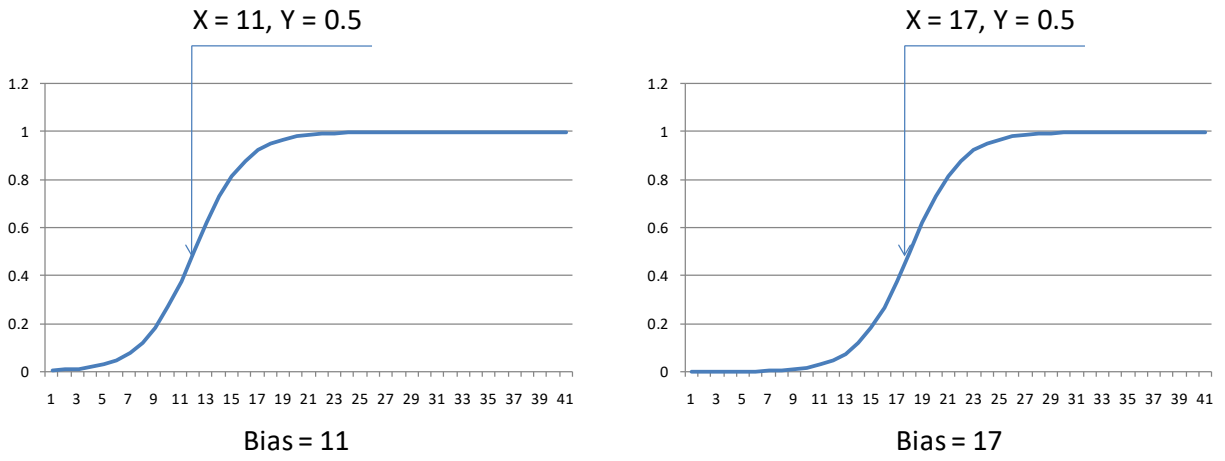


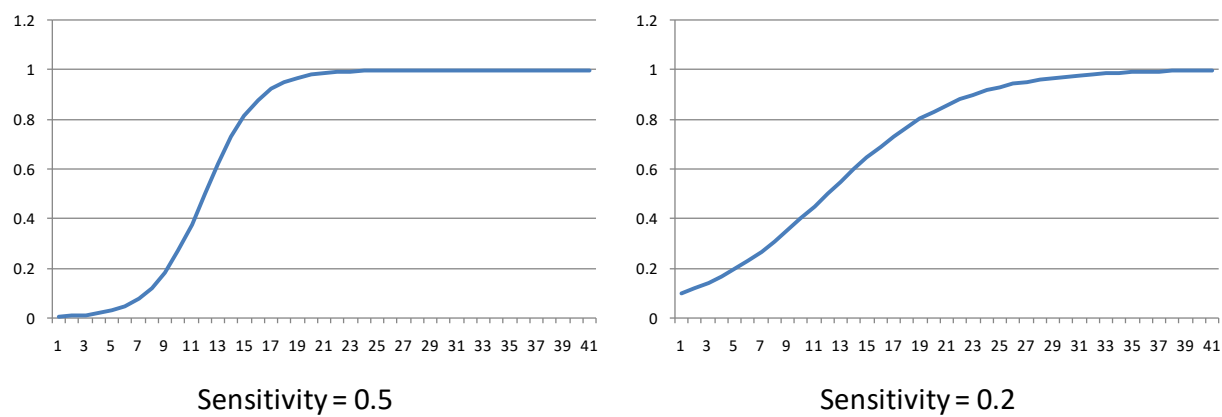
Table 1 Input Data Matrix

Attribute	Item																																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	
a_1	1	2	3	4	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
a_2	0	0	0	0	0	1	2	3	4	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
a_3	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	3	3	3	3	6	6	6	6	6	6	6	6	6	6	6
a_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	1	2	3	4	2	2	2	2	2	2	2
a_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6

Figure 31: Kohonen’s SOM Diagram

**Figure 32: MSE on Sigmoid**

**Figure 33: Bias Parameter**

**Figure 34: Sensitivity Parameter**

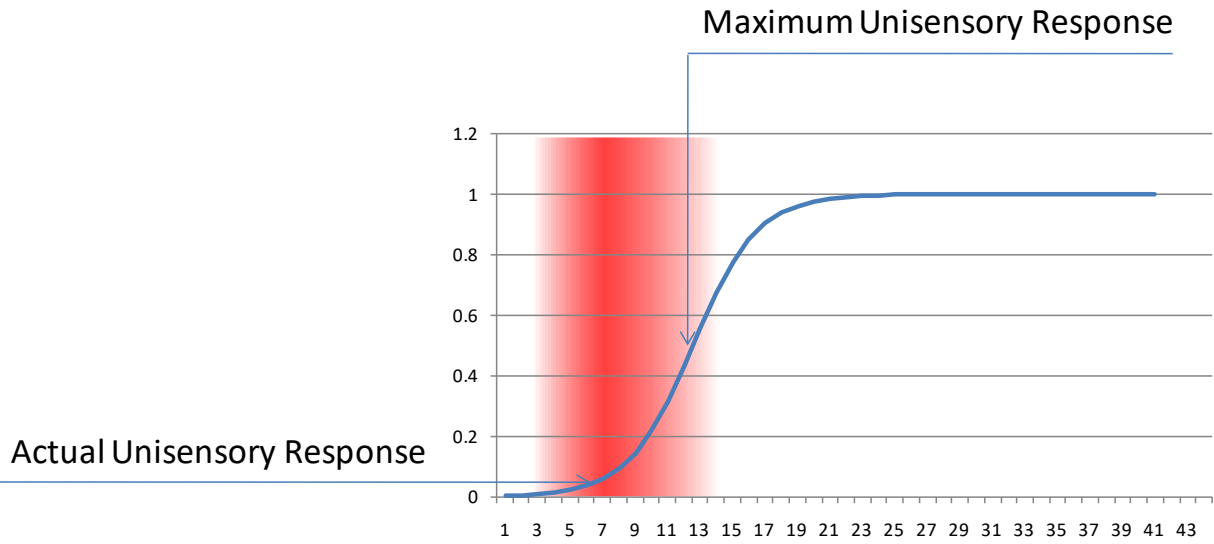


Figure 35: MMA Sigmoidal Tuning

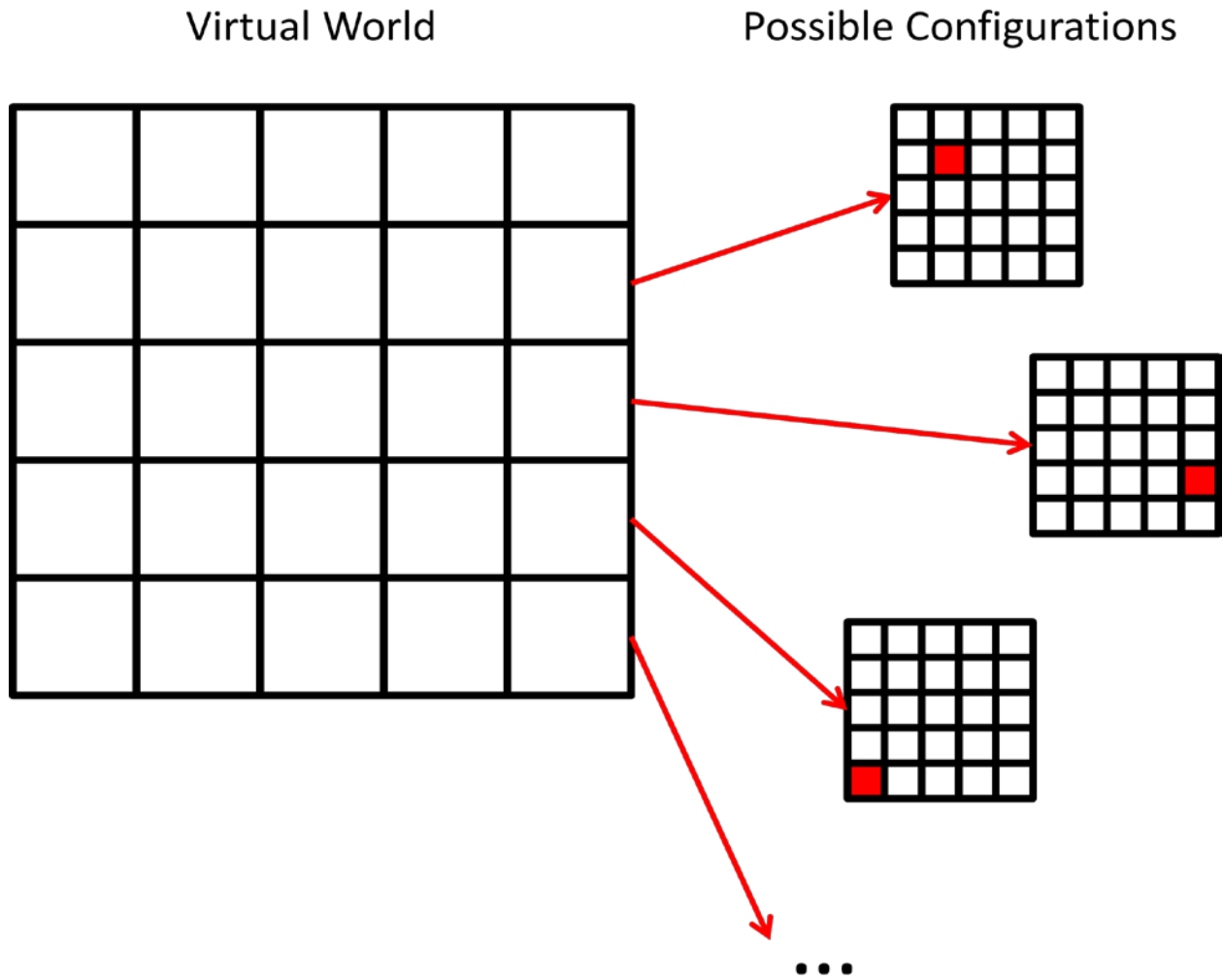


Figure 36: Virtual World Generation

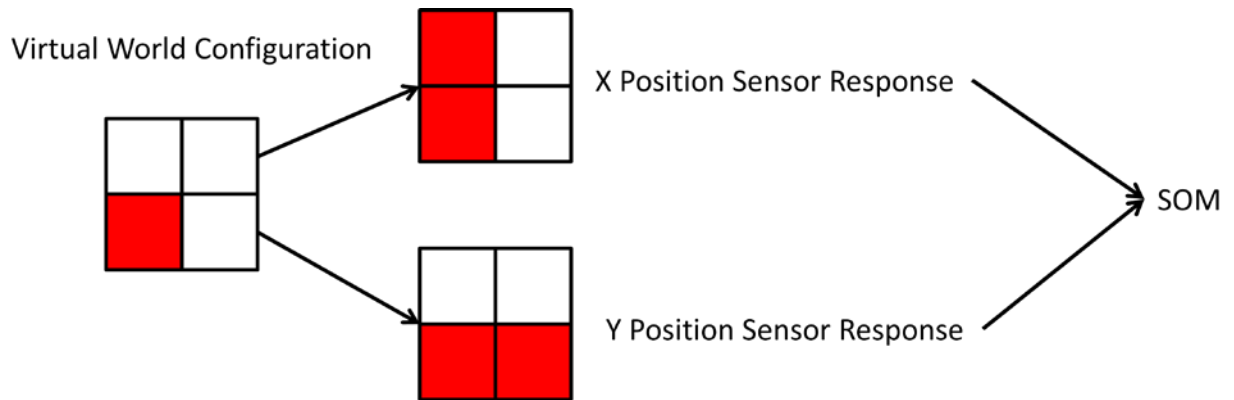


Figure 37: X Position and Y Position Sensors

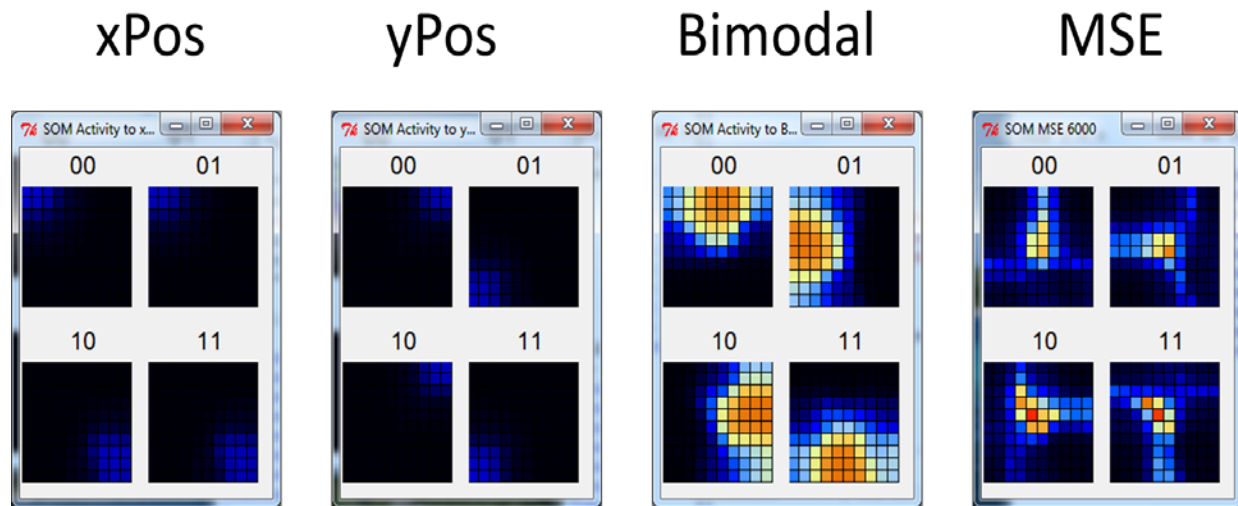


Figure 38: Results from xPos and yPos Trial with MMA Parameters

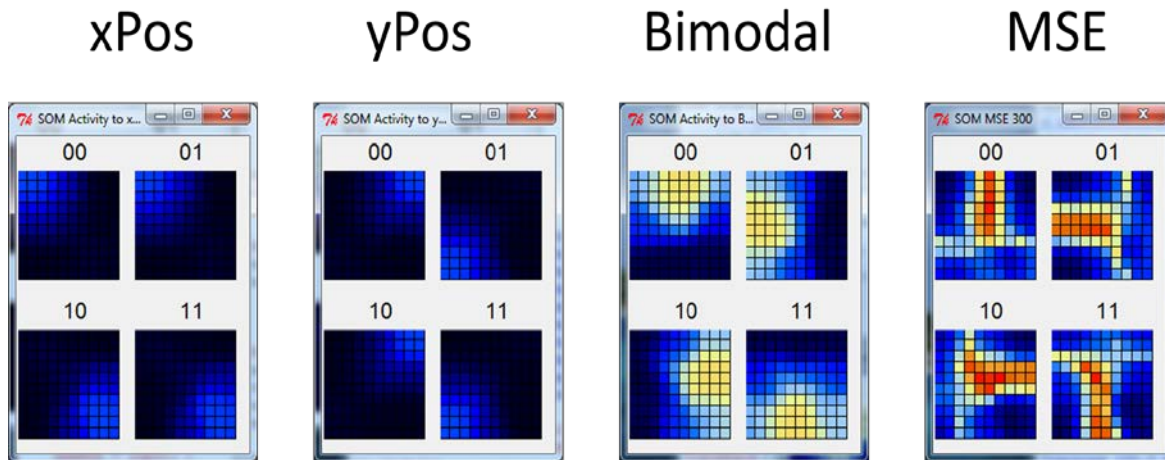


Figure 39: Results from xPos and yPos Trial with Modified Sensitivity

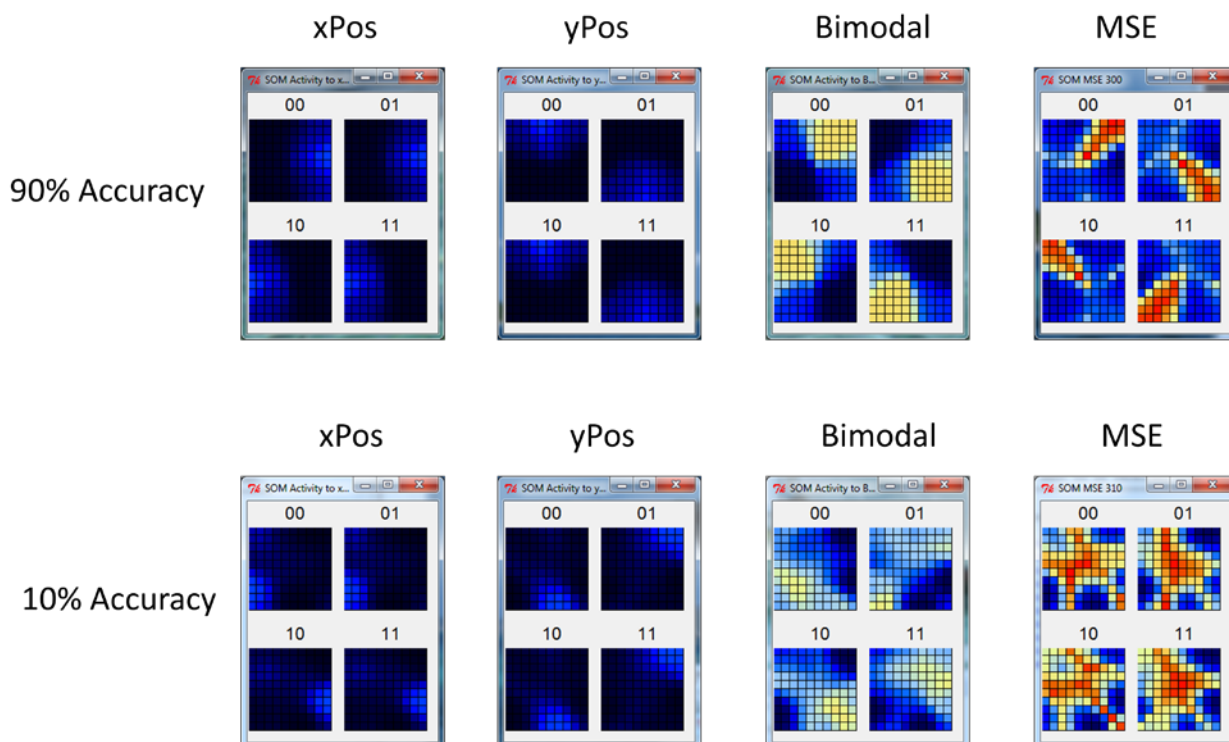


Figure 40: Results from xPos and yPos Trial with Modified Accuracy

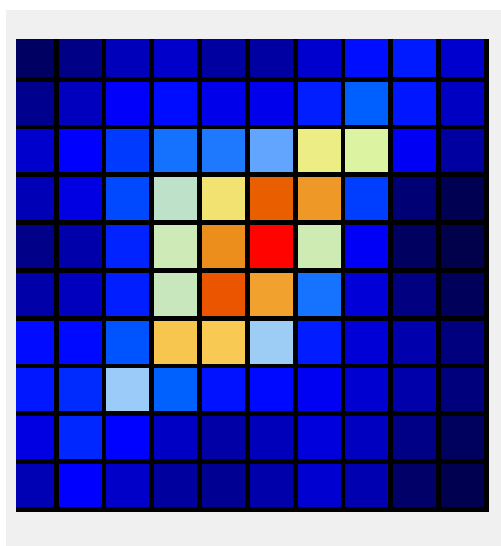


Figure 41: MSE in Response to Two Unimodal Stimuli in the Same Sense

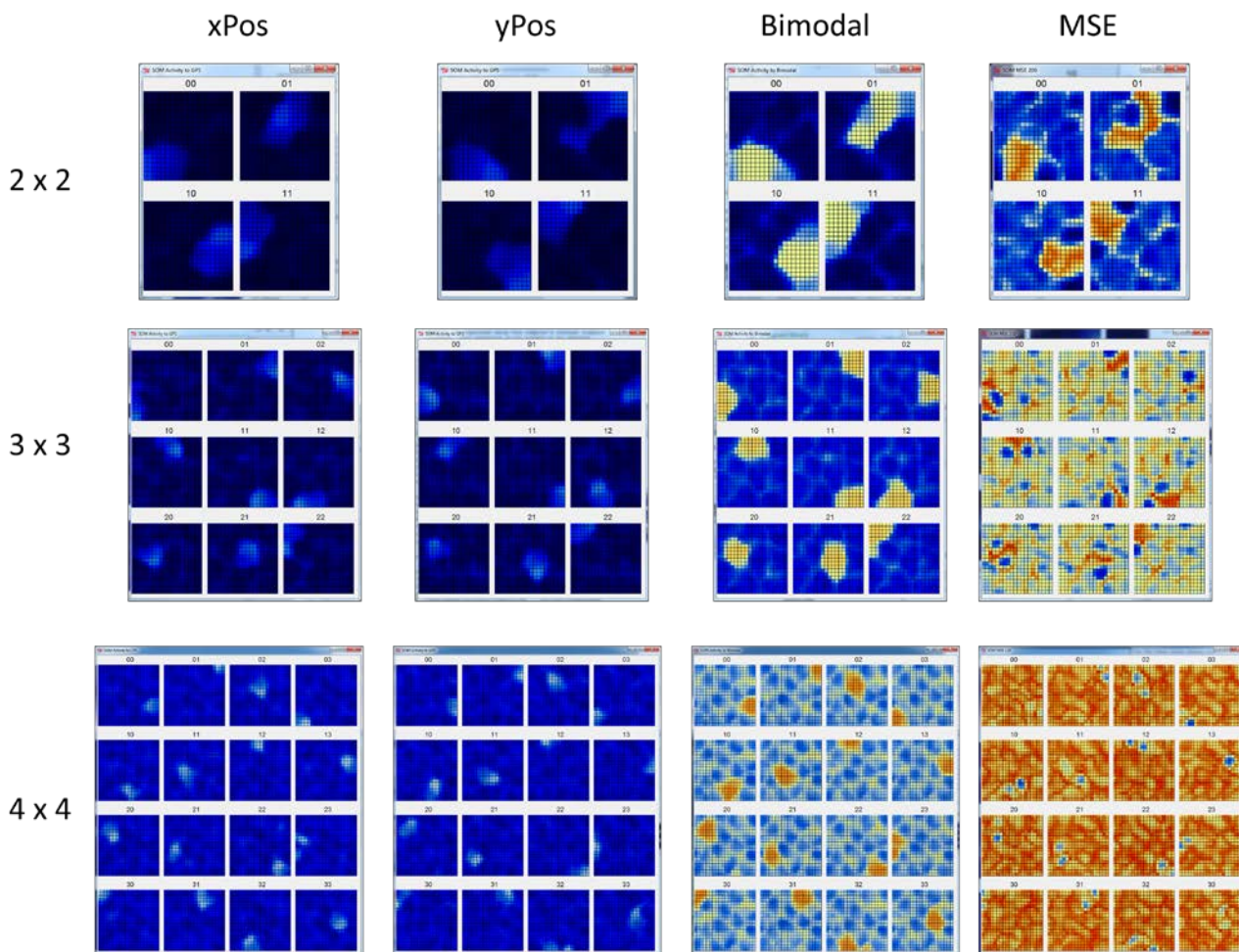


Figure 42: Increasing Grid Size with MMA Parameters

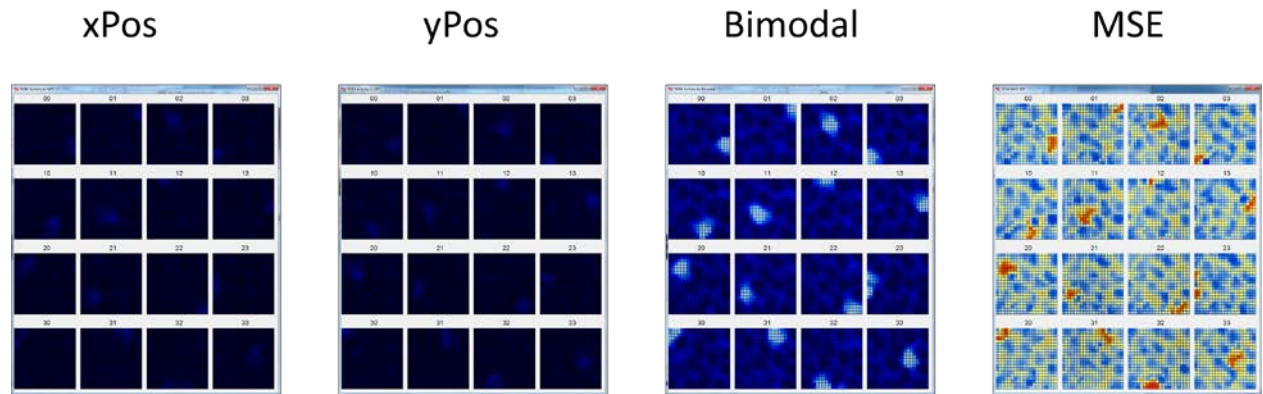


Figure 43: 4x4 Grid With Adjusted Bias

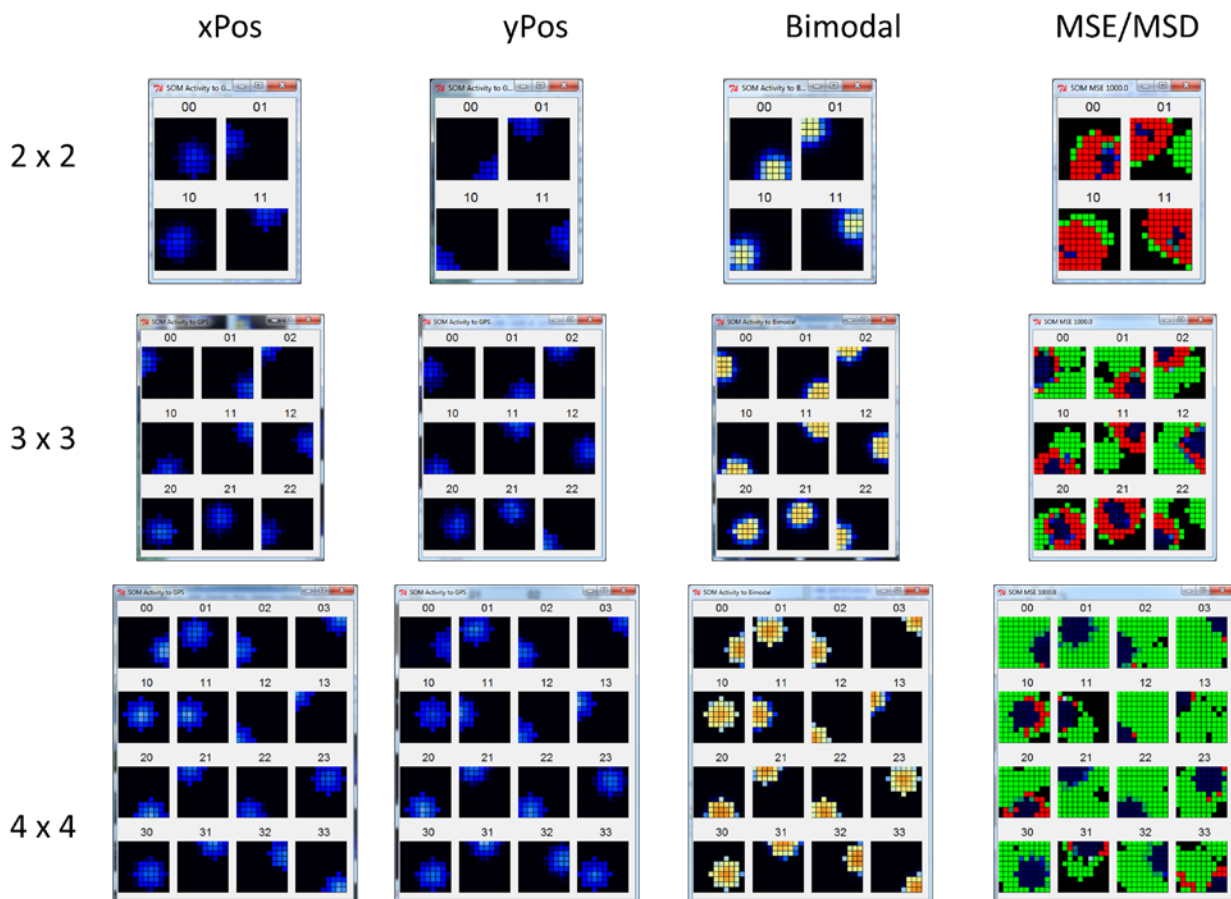


Figure 44: MSE and MSD for Inhibition Extension

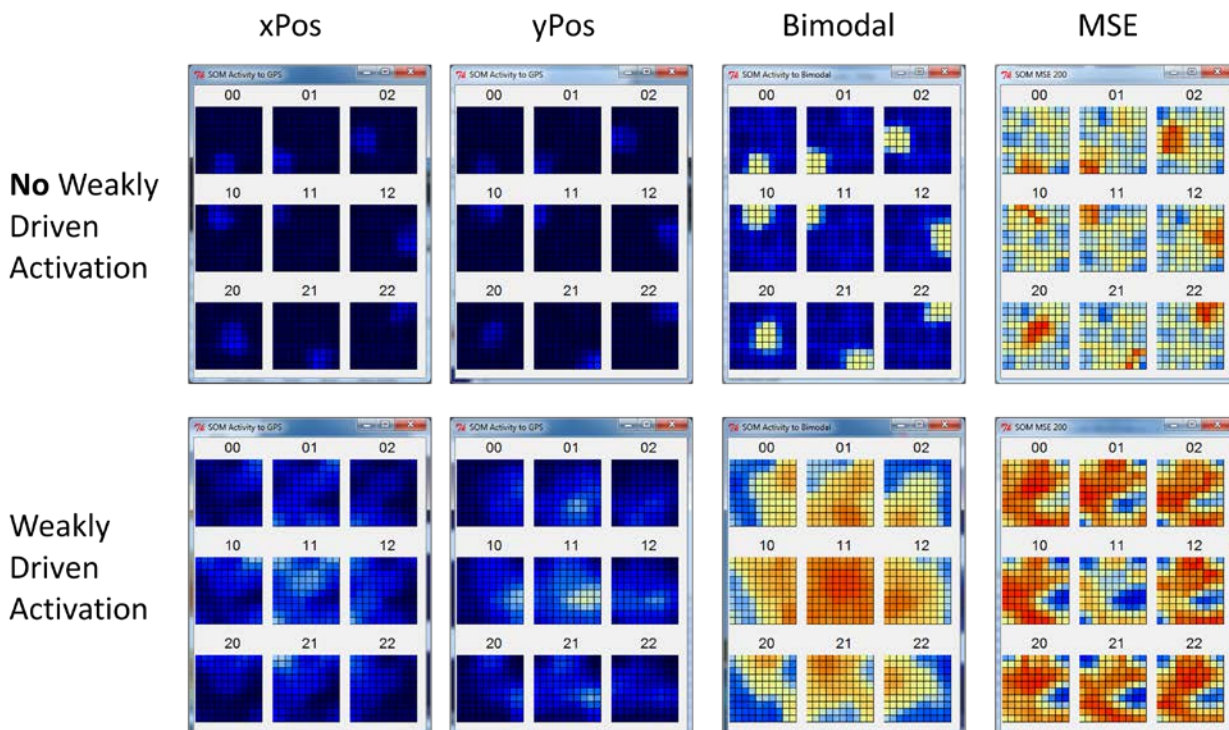


Figure 45: MSE With and Without Weakly Driven in 3x3

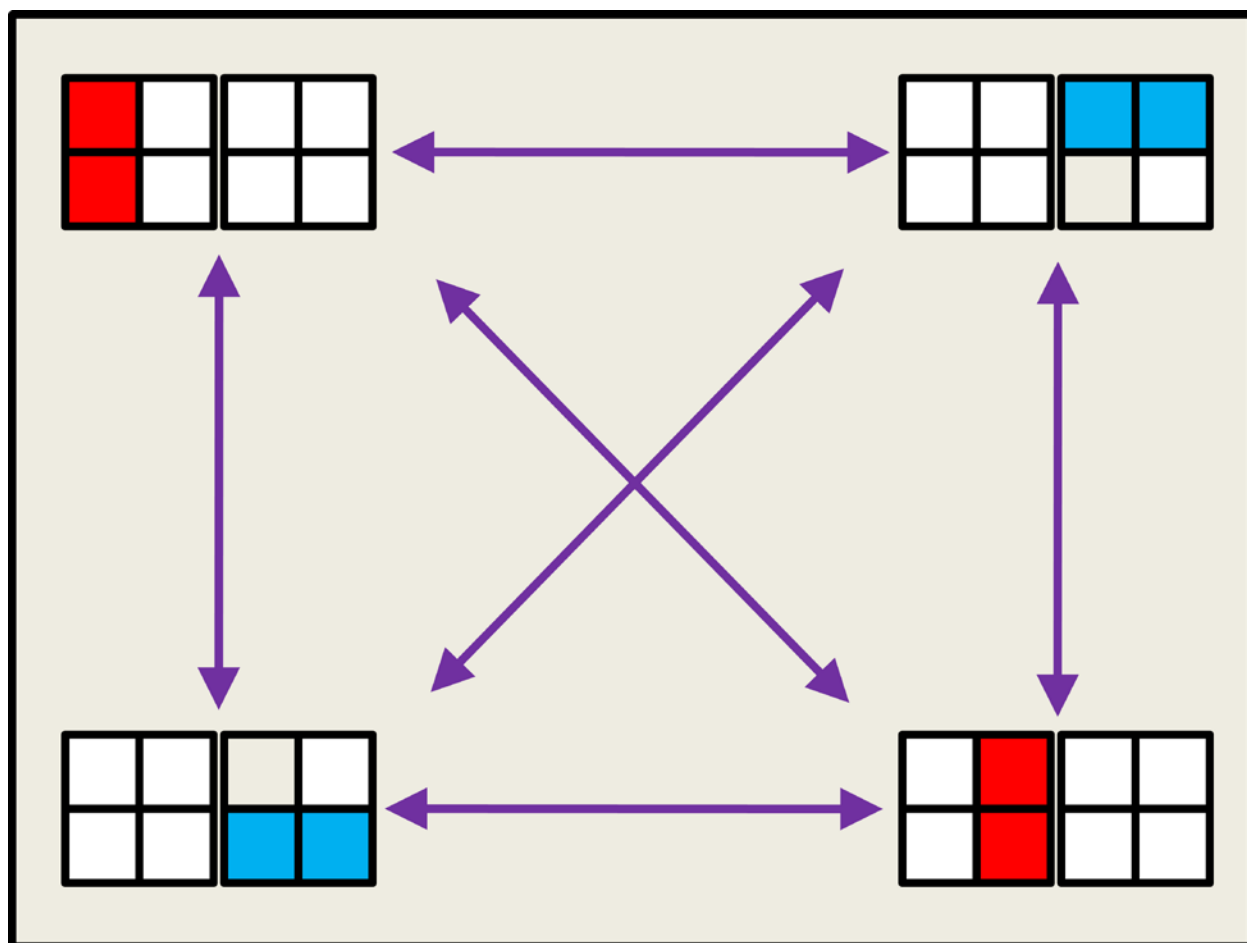


Figure 46: Simplified Weight Distribution

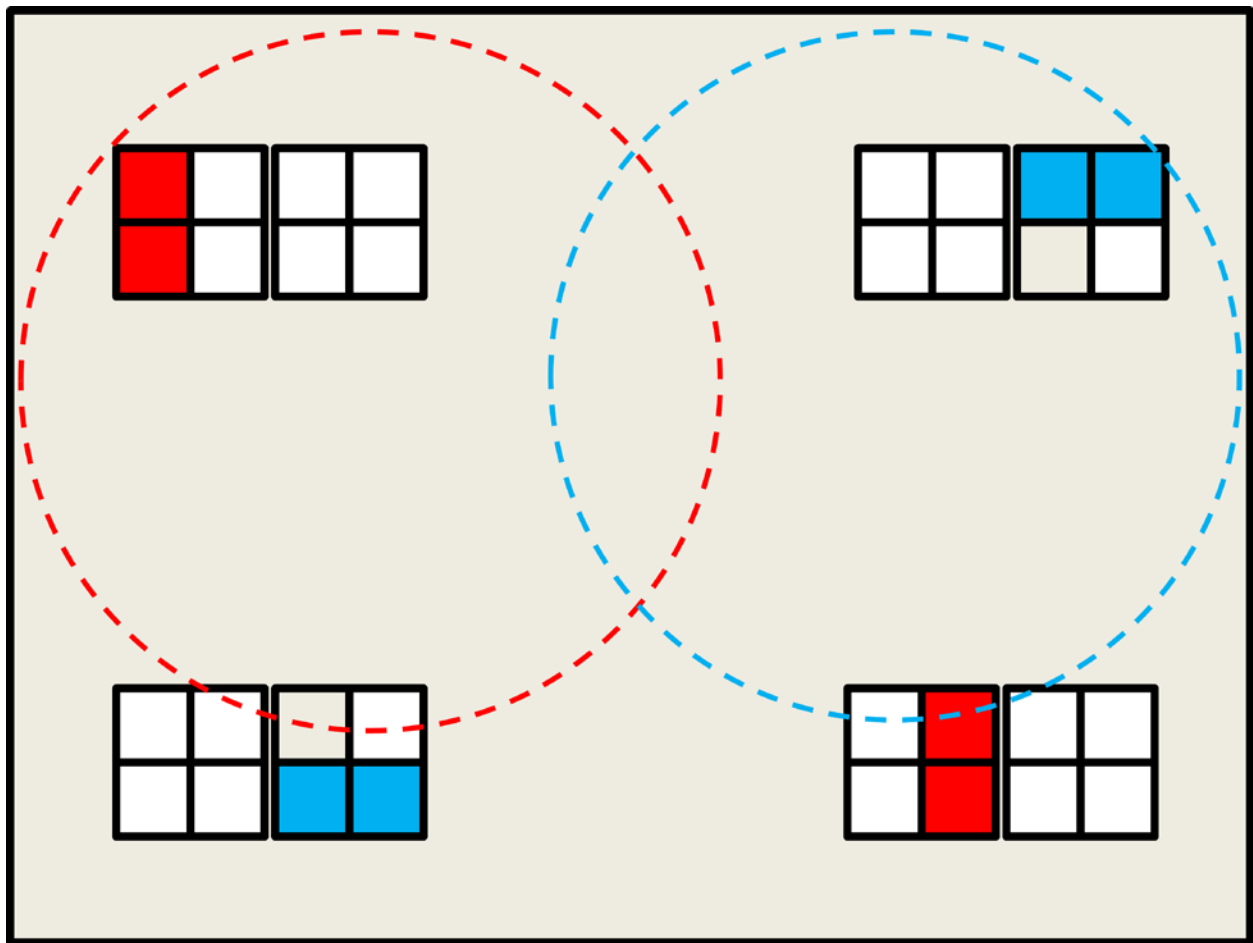


Figure 47: Overlapping Neighborhoods Between Senses

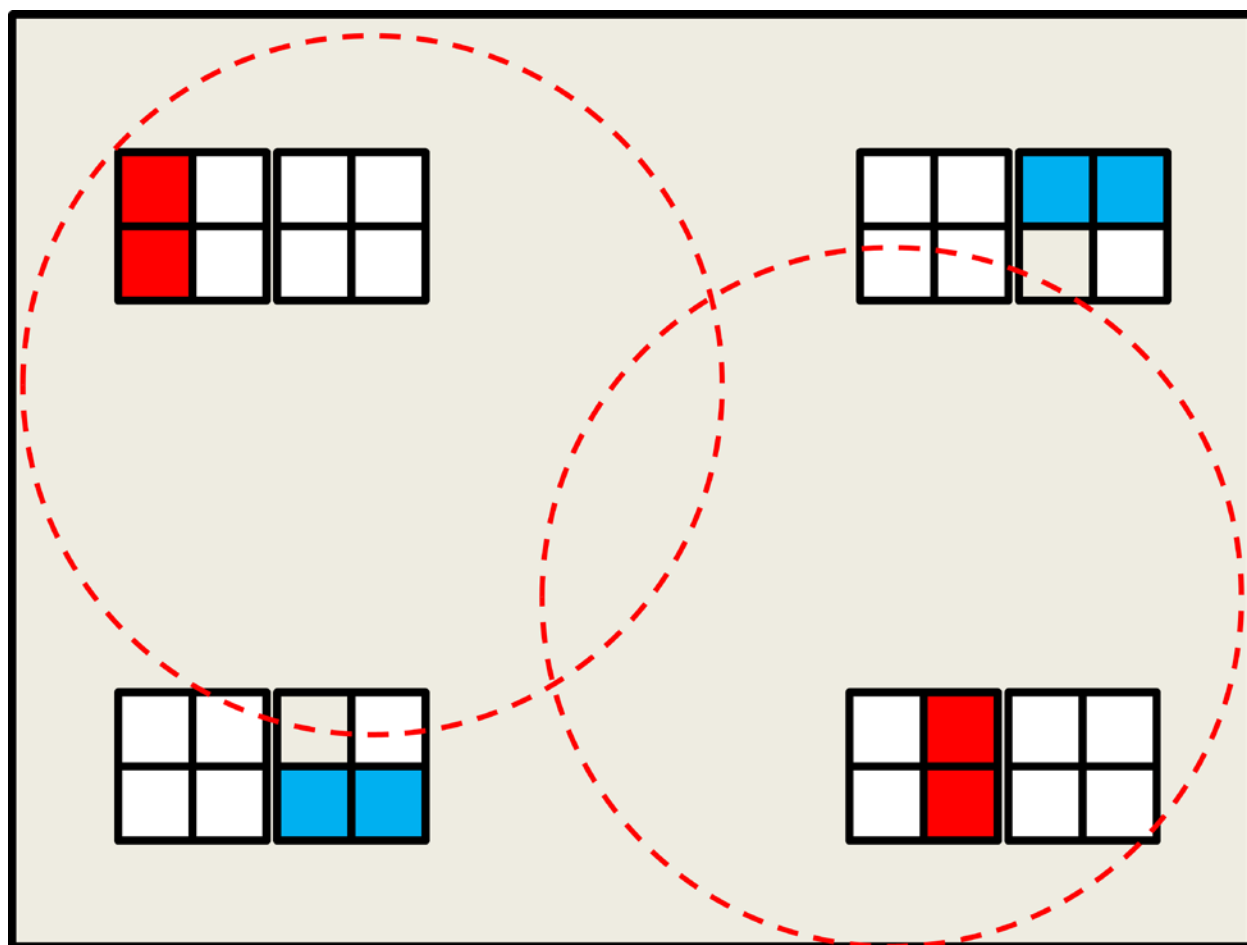


Figure 48: Overlapping Neighborhoods Within a Sense

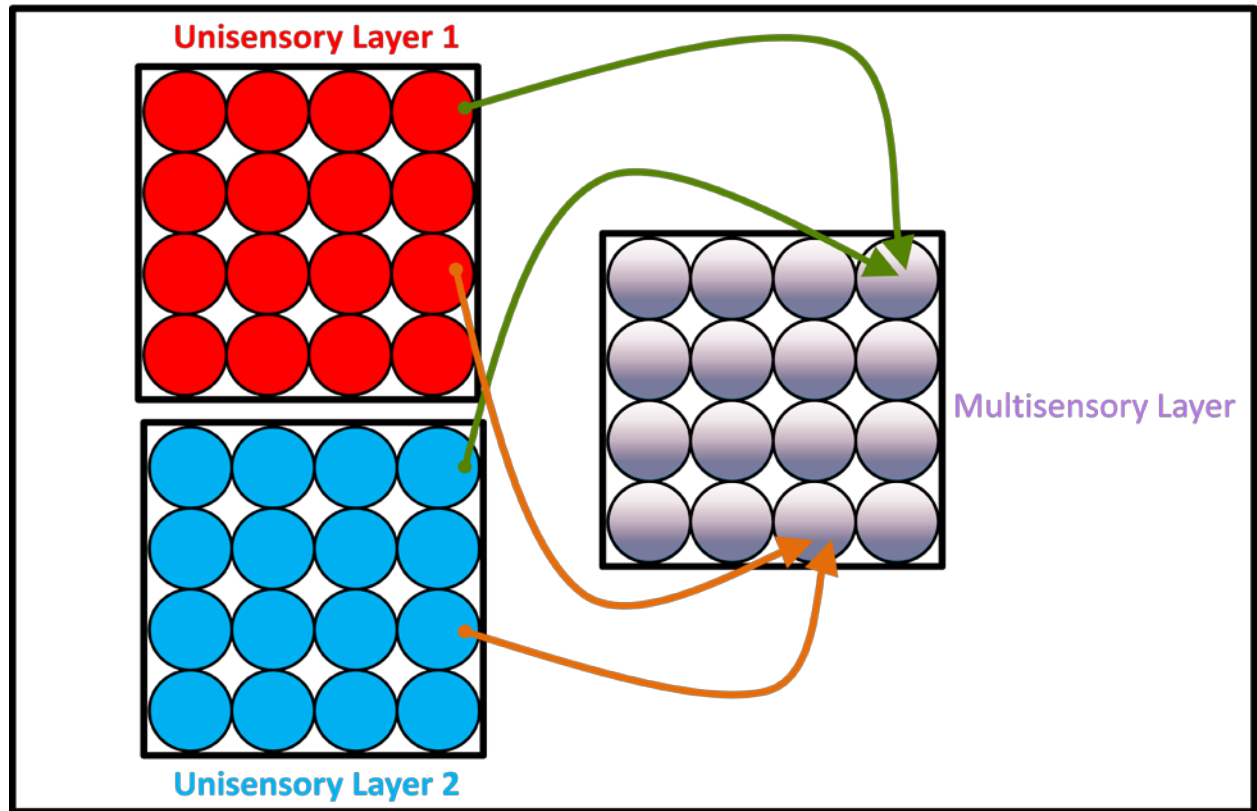


Figure 49: Diagram of Theoretical Model

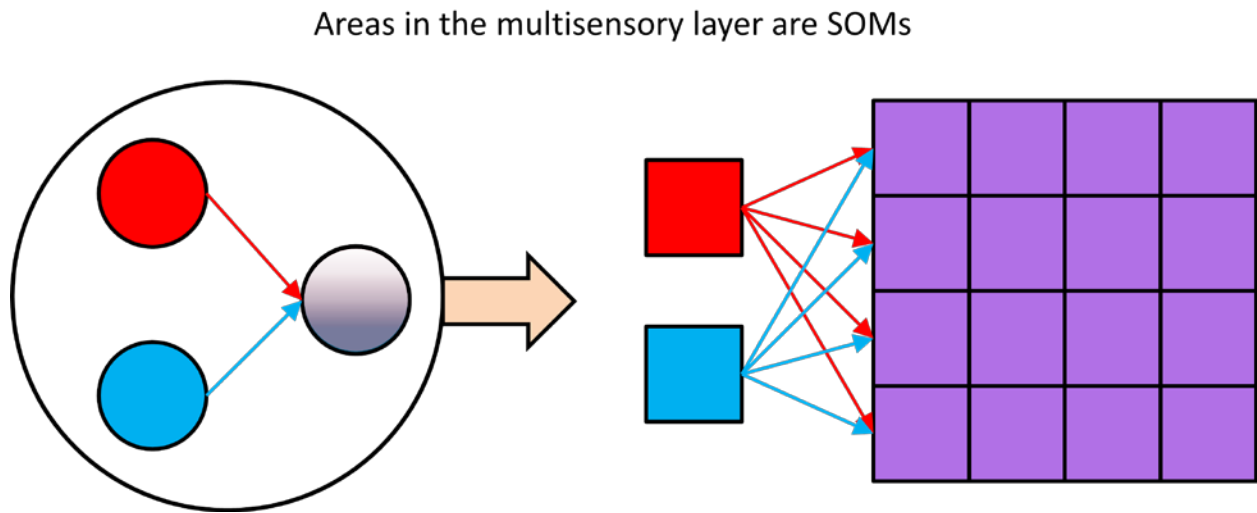


Figure 50: Diagram of Projections of Areas within Layers

References

- Alvarado, J. C., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2008). A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Research*, 1242, 13-23.
- Anastasio, T. J., & Patton, P. E. (2003). A two-stage unsupervised learning algorithm reproduces multisensory enhancement in a neural network model of the corticotectal system. *Journal of Neuroscience*, 23, 6713-6727.
- Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*, 7, 541-546.
- Bavelier, D., Dye, M. W. G., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, 10, 512-518.
- Bienenstock, E., Cooper, L. N. & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interactions in visual cortex. *Journal of Neuroscience* 2:32-48.
- Brett-Green, B. A., Miller, L. J., Gavin, W. J., & Davies, P. L. (2008). Multisensory integration in children: A preliminary ERP study. *Brain Research*, 1242, 283-290.
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Research*, 1242, 126-135.
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57, 11-23.
- Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., & Stein, B. E., 1979 "Sources of succortical projections to the superior colliculus in the cat" *Journal of Comparative Neurology* 184 309-329.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162-169.
- Falchier, A., Clavagnier, S., Pascal, B., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22, 5749-5759.
- Foxe, J. J. (2008). Toward the end of a "principled" era in multisensory science. *Experimental Brain Research*, 1242, 1-3.
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10, 77-83.

- Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, 16, 419-423.
- Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, 25, 5004-5012.
- Gori M. Del Viva, M., Sandini, G., & Burr, D.C., 2008. Young children do not integrate visual and haptic form information. *Current Biology*.18, 694-698.
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10, 278-285. .
- Jordan, K. E., Suanda, S. H., & Brannon, E. M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition*, 108, 210-221.
- Killackey, H.P., Rhoades, R. W. & Bennett-Clarke, C.A. (1995). The formation of a cortical somatotopic map. *Trends in Neuroscience*. 18, 402-407.
- Kim, J., & Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Research*, 1242, 263-275.
- Kohonen T. (1990). The self-organizing map. *Proc Inst Electrical Electronics Eng*, 78, 1464-80.
- Kohonen, T. & Hari, R. (1999). Where the abstract feature maps of the brain might come from. *Trends in Neuroscience*. 22, 135-139.
- Martin, J. G., Meredith, M.A. & Ahmad, K. (2009). Modeling multisensory enhancement with self-organizing maps. *Frontiers in Computation Neuroscience*, 3.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, 14, 115-128.
- Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, 1242, 191-199.
- Philippi, T. G., van Erp, J. B. F., & Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain Research*, 1242, 116-125.
- Rauschecker, J. (1999). Auditory cortical plasticity: A comparison with other sensory systems. *Trends in Neuroscience*, 22, 74-80.
- Robert, M. (1973). Cross-cultural research in the perception of pictorial materials. *Psychological Bulletin*, 80, 135-150.

- Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International journal of Psychophysiology*, 50, 19-26.
- Rowland, B. A., Stanford, T. R., & Stein, B. E. (2007). A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception*, 36, 1431-1443.
- Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology*, 15, 454-458.
- Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, 85, 1322-1327.
- Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology*, 50, 5-17.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12, 411-417.
- Small, D. M. (2004). Crossmodal integration – insights from the chemical senses. *Trends in Neurosciences*, 27, 120-123.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9, 255-266.
- Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, 198, 113-126.
- Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, 194, 91-102.
- Ursino, M., Cuppini, C., Magosso, E., Serino, A. & Pellegrino, g. (2009). Multisensory integration in the superior colliculus: a neural network model. *Journal of Computational Neuroscience*, 26, 55-73.
- Wagner, H., Kautz, D. & Poganiatz, I. (1997). Principles of acoustic motion detection in animals and man. *Trends in Neuroscience*. 20, 583-588.
- Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience*, 17, 2429-2444.