Capacity limits in oscillatory networks: implications for sensory coding

A. Ravishankar Rao, Guillermo Cecchi

Abstract—Psychological studies have investigated limits in the capacity to simultaneously store multiple objects in working memory, which turns out to be approximately four. In this paper we examine the existence and origin of such a capacity limit in the sensory encoding stage, where synchronous activity can be considered to group related features of a common object. We develop a model of an object recognition network using oscillatory elements that can achieve phase synchronization.

Using simulations based on this network, we show that distinct phases of oscillation can be used to label combinations of objects presented simultaneously. This allows the network to separate mixtures of objects, and identify the input elements that belong to each object. We demonstrate that there is a limit of four objects that can be separated from a mixture. Further studies are required to generalize this result by varying the size of the network and the number of objects used for training. We also show that by narrowing a tuning function that governs the dynamics of the system, we can achieve higher separation accuracy. However, this comes at the cost of utilizing a higher number of iterations over which the system settles and learns its synaptic weights. We lay down a framework for the quantitative modeling of factors affecting capacity limits, which has the potential to advance our understanding of sensory representation, attention, working memory and multitasking.

I. INTRODUCTION

The work of the psychologist Miller [1] in his paper entitled “The magical number seven plus or minus two”, suggested that there may be fundamental limits on the information processing capability of the human brain. It is intriguing that though the human brain contains billions of neurons we are capable of attending to only a few thoughts or stimuli simultaneously. An early psychological study by Sperling [2] investigated the ability of subjects to recall an array of up to 12 characters flashed briefly on a screen. Sperling found that subjects could recall only about 4 characters on average.

Spurred by these observations, Cowan [3] proposed a schematic representation of working memory depicted in Figure 1. The sensory input is contained in a sensory store, which is conveyed to the activated memory. Only certain objects are recognized by the focus of attention, which has a limited capacity. The notion of capacity limits in working memory are derived based on the inherent capacity limits of the focus of attention shown.

Though such a schematic has proven to be very useful in the psychology literature, it does not provide a mechanistic view of how the neural processing occurs at the lowest levels of single neuron activity. The theory and experimentation in psychology is concerned with organism level behavior, and not necessarily with specific neural substrates that produce the behavior.

In the current paper, we develop the viewpoint that capacity limits may exist as early as the sensory coding stages. We aim to bridge the gap between the behavior of lower level units in the neural system and observed higher level characteristics such as capacity limits.

Our work is based on the phenomenon of synchronization that has been observed in neural circuits [4]. Salinas and Sejnowski [5] present a review of research showing that attention helps to synchronize activity in the visual and somatosensory cortices. This mechanism of synchronization serves to route input signals to higher level cortical areas.

Our earlier work [6], [7] presented a model of an oscillatory network that achieved feature binding through phase synchronization over multiple layers of a network. We suggested that the phase of oscillations could form a tag of relatedness of features representing a given object across these multiple layers.

There appears to be increasing evidence from the neuroscience field indicating that synchrony helps to coordinate and relate neural activity in a distributed network. Akam and Kullman [8] investigated a mixture of synchronous and asynchronous channels, and showed that information can be routed successfully through synchronous channels. Recently, Buschman et al. [9] showed that ensembles of synchronized activity in the prefrontal cortex represent distinct sets of rules. They refer to the ability of synchronized oscillations to dynamically ‘carve out’ ensembles of neurons from a large population.

When we juxtapose the findings of capacity limits with the observations of synchronous activity leading to ensembles of grouped information, this leads to a fundamental question: is there a relationship between the representation of...
synchronous activity and a capacity limit? In other words, if multiple phases of oscillation represent multiple related objects, is there a limit to how many objects can co-exist in the phase space?

Figure 2 illustrates this concept. We have shown an idealized case with two objects grouped by phase similarity. The question we raise is: how many objects can be simultaneously maintained with different phase synchronizations if the system has to perform a reliable distinction between these objects?

![Fig. 2. An idealized representation of oscillators in a two-level hierarchical neural system, labeled Layer 1 and Layer 2 which project to each other. The phases of oscillators are represented by the direction of the arrows within the ellipses. Synchronized oscillators having the same phase represent a tag of relatedness, and may be considered to group features corresponding to a single object. We depict two objects here with distinct phases. In Layer 1, the first object is a triangle, represented in blue, where the elements of the triangle have the same phase. The second object is a square, represented in green, where the elements of the square have the same phase, but different from the phase of the triangle. The rest of the units in layer 1 are not synchronized with either of these two objects. In Layer 2, there are units that are synchronized with the phases of the square and triangle in Layer 1, and also other units that are not synchronized. Thus, we can see that the phase of synchronization serves as a tag that allows units representing a common percept to be identified, or 'carved out' at multiple levels of the hierarchy.](image)

Theoretically, there should be no limit, as the phase is a continuous variable in the range \([0, 2\pi]\), and multiple objects could possess multiple phases. However, if there is a behavioral task involved, such as the generation of a response to a certain object with a given phase, then it is possible to be distracted by objects with similar phases, which causes the task performance to deteriorate.

The research presented in this paper probes the following questions.

- How many objects can be simultaneously represented via the phase of oscillations in a sensory stream without affecting the ability to distinguish them?
- Can we identify low-level features related to oscillatory activity that may be able to predict higher-level system behavior?

We build on our earlier work that utilized an objective function incorporating sparse spatio-temporal encoding to derive the oscillatory dynamics of a network. The novel contribution of the current paper lies in our use of simulations to show that up to four objects can be simultaneously represented before the ability to distinguish them reaches a chance level. We also show that the shape of the probability density function of phase angles is highly correlated with the ability of the network to separate multiple objects. We use the kurtosis as a measure of the flatness of the phase distribution, and show that higher kurtosis values are associated with a greater ability to separate objects.

The remainder of this paper is organized as follows. Section II describes the oscillatory neural network used in our simulations. Section III explains our experimental findings. In Section IV we explore the implications of our findings.

II. METHODS

We briefly summarize the computations used in the model for the two-layer system of oscillators as shown in Figure 2. The system has been presented in detail in our earlier paper [6], and we review only the key elements. Let \(x\) denote units in the lower layer and \(y\) denote units in the upper layer, which are connected by a weight matrix \(W\). Each unit is considered to be an oscillator with an amplitude, frequency and phase of oscillation. If all the units have a similar nominal frequency, their behavior can be described in terms of phasors of the form \(x_n e^{i\phi_n}\) for the lower layer and \(y_n e^{i\theta_n}\) for the upper layer. Here, \(x_n\) and \(y_n\) represent amplitudes of the \(n^{th}\) unit in the lower layer and upper layer respectively, and \(\phi_n\) and \(\theta_n\) are their corresponding phases.

\[
\begin{align*}
\Delta y_n & \propto \sum_j W_{nj} x_j \left[ 1 + \cos(\phi_j - \theta_n) \right] - \alpha y_n \\
& - \gamma \sum_k y_k \left[ 1 + \cos(\theta_k - \theta_n) \right] \\
\Delta \theta_n & \propto \sum_j W_{nj} x_j \sin(\phi_j - \theta_n) \\
& - \gamma \sum_k y_k \sin(\theta_k - \theta_n) \\
\Delta \phi_n & \propto \sum_j W_{nj} y_j \sin(\theta_j - \phi_n)
\end{align*}
\]

Given a set of initial conditions, Equations 1 - 3 describe the instantaneous evolution of the system. We assume the lower layer is initialized to pixel values of a 2-D visual image representing a visual stimulus. We choose this interpretation simply to make the functioning of the system easy to understand. In general, the lower layer could represent any cortical area. The initial values for the upper layer, \(y\) are set to zero. The values of the phases are randomized.

The update rules in Equations 1 - 3 are applied, causing the system to exhibits transients and eventually settle down, say after 200 iterations. The synaptic weights \(W\) are then modified as follows.

\[
\Delta W_{ij} \propto y_i x_j \left[ 1 + \cos(\phi_j - \theta_i) \right]
\]

In our simulation, the lower layer consists of 8x8 units, each of which receives an image intensity value as input. The upper layer \(y\) consists of 16 units. There are all-to-all
connections between units in the lower layer $x$ and the upper layer $y$. The units in the upper layer possess all-to-all lateral connections. Finally, there are all-to-all feedback connections from $y$ to $x$. After learning, we observe a winner-take-all dynamics in the network when it is presented with one of the learned inputs.

We choose an input set consisting of 16 simple visual objects such as a square, triangle, cross, circle and so on, some of which are shown in Figure 3. The complete set can be found in [6].

The network operates in two stages consisting of learning and performance. During the learning stage, we present a randomly selected object as input, and we let the network activity settle. Following this, we apply the Hebbian learning rule of equation 4. We repeat this process over 1,000 trials. The system typically shows a winner-take-all behavior at the upper layer $y$ for each input presented. Furthermore, after the training, a unique winner is associated with each input. Note that the network training proceeds in an unsupervised fashion.

As shown in [6], when two inputs are combined and presented to the lower layer $x$, it results in two units, termed the winners, being activated in the upper layer $y$. (Strictly speaking, a winner implies a single unit. However, we consider the top two units with the highest amplitudes, and use the term winner for brevity, rather than identifying a winner and a runner-up explicitly). These units are the ones with the two highest amplitudes. Furthermore, the phases of these two winners in layer $y$ are synchronized with the phases of units in the lower layer $x$ that correspond to the two individual inputs. The interpretation of this behavior is that different units can be simultaneously active while having phases that are maximally apart from each other [7].

We define a measure termed the separation accuracy, which captures the ability of the network to correctly identify mixtures of inputs. Suppose unit $i$ in the upper layer is the winner for an input $x_1$, and unit $j$ is the winner for input $x_2$. If units $i$ and $j$ in the upper layer are also winners when the input presented is the mixture $x_1 + x_2$, then we say the separation is performed correctly, otherwise not. The ratio of the total number of correctly separated cases to the total number of cases investigated is the separation accuracy. Similarly, this concept can be extended to mixtures of multiple objects, and the separation accuracy reflects how accurately the network is able to correctly identify the components of the mixture.

A novel contribution of this paper is to investigate the relationship between the separation accuracy and the number of objects combined in the mixture, as shown in Section III.

III. RESULTS

Figure 4 shows the variation of separation accuracy with respect to the number of objects simultaneously presented. We compare the separation accuracy against a chance accuracy calculated as follows. If the system is told that $k$ objects were presented, and it was randomly guessing the identity of these objects, the chance accuracy is $1/C(n,k)$ where $C(n,k)$ is the number of $k$-combinations from the set of $n$ elements. In our case, $n=16$, and $k$ varies from 1 to 4. The separation accuracy of the system is well above chance for up to 3 simultaneously presented objects. The error-bars for the separation accuracy are shown for each case. The interesting aspect of this figure is that for 4 objects, the error-bar begins to hit the chance level accuracy. This indicates that the system is not able to reliably identify the constituents of a mixture of 4 objects presented to the input layer $x$. This may be viewed as a capacity limit on the number of objects that can be separated by such a system. To the best of our knowledge, this is the first quantitative demonstration of a capacity limit in oscillatory networks where phase is used to identify objects.

We now investigate whether there is a relationship between the distribution of object phases and the capacity limit. In the case of two objects, the phases of the winners representing these objects are maximally apart at 180 degrees, resulting in low ambiguity in discriminating them. The intuition is that as the number of objects increases, the phase differences between units representing these objects decreases, and this may result in poorer system performance.
In Figures 5 - 8 we examine the distribution of phase angle differences in the two-layer system as the number of simultaneously present objects increases. We select the smallest phase angle as the reference phase, and compute the phase differences with respect to the reference phase.

We start with a presentation of two randomly selected objects from the input set, and increase the number of objects to four.

As can be seen from Figure 8, the probability density function of phase differences gets flattened out as the number of objects increases. This confirms the intuition we laid out earlier, as we expect the system will have difficulty in separating phases as the number of objects increases.

In order to quantify the notion of the flatness of a distribution, we compute the kurtosis of the distributions shown in Figure 8. The kurtosis measure of a peaked distribution is greater than that of a flat distribution. The kurtosis is shown in Figure 9 and captures the flattening of the phase difference probability density function as the number of objects is increased.

The correlation between the kurtosis values in Figure 9 and the separation accuracy in Figure 4 is 0.96 (pval = 0.189). This high correlation value indicates that the probability density function of phase differences in the neural system is a reasonable substrate to explain the separation accuracy. Thus, we are able to tie a lower level attribute, the kurtosis of a phase difference distribution, with a higher-level attribute, the system performance in separating mixtures of inputs. This analysis provides a mechanistic interpretation of what may be a neural correlate of a capacity limit in the sensory stream.
A. Varying the tuning function

Based on Equation 1, the upper layer amplitudes $y_n$ depend on the phase differences between the input unit $j$ in the lower layer, and unit $n$ in the upper layer as follows.

$$\Delta y_n \propto [1 + \cos(\phi_j - \theta_n)] \quad (5)$$

This is plotted in Figure 10(A), and may be considered to be a tuning function that determines the effect of a given phase difference $\phi_j - \theta_n$ on the amplitude, $y_n$. This tuning function can be made narrower as shown in Figure 10(B).

We examine the effect of the tuning function on the network performance. The intuition here is that a narrow tuning function will allow more discrete phases to be represented in phase space, and this could possibly increase the capacity limit.

Figure 11(A) shows the performance of the network in terms of separation accuracy as we increase the number of objects simultaneously presented. The narrower tuning function in Figure 10(B) is employed, and the number of settling iterations used is 200. We see that the performance of the system is not as good as the one shown in Figure 4.

However, as the number of settling iterations is increased to 400, the performance of the system improves, as shown in Figure 11(B).

In Figure 11(C), we superpose the plot in Figure 11(B) on the plot from Figure 4.

These results indicate that there is a tradeoff between the width of the tuning function and the number of settling iterations. As the tuning function is made narrow, and the number of settling iterations is increased, the accuracy of simultaneous object representation can be increased.

IV. DISCUSSION

We wish to point out that it is quite possible to design a fixed network that can separate mixtures of known objects reliably. This can be done by precisely selecting the weights a-priori to achieve optimal phase separation for each object. However, there is no learning that such a system can exhibit, and it will work only for the specific objects it has been programmed to identify. In contrast, our system performs unsupervised learning, and can learn its weights to accommodate any set of input objects. This is a desirable attribute if the goal is to develop an understanding of the neural mechanisms that may be involved in object recognition and binding.

Cowan [10] describes two perspectives to approach the issue of capacity limits. These are the processing-related and storage-specific perspectives. Working memory capacity is influenced by the specific process applied to a task, such as verbalizing items to be remembered, or forming a sequence of visual imagery associated with the items. The research presented in the current paper would be considered to be a processing-related limit on the sensory stream. Cowan [10] offers the conjecture that “if the neural patterns for multiple concepts are instead active concurrently, it may be that more than about four concepts result in interference among them, or that separate brain mechanisms are assigned to each concept, with insufficient neurons at some critical locale to keep more than about four items active at once.”

Our work provides a computationally grounded perspective that substantiates this conjecture. We do observe that the presence of four visual objects causes an interference in the ability to distinguish them. However, since there are a sufficient number of units in our upper layer, which is 16 in our simulation, it appears that the capacity limit is not due to insufficient units being available. Our work suggests that there is a more fundamental limitation that is at play.

Lisman and Idiart [11] proposed a model for short term memory where oscillations in two frequency bands consisting of the theta (5-10 Hz) and gamma (20-60 Hz). Memory capacity is determined by how many fast cycles can occur within the slower cycle. Our current model does not consider multiple frequencies. We have been able to show that a capacity limit can exist even within a single frequency band. It is possible that our approach could be extended to consider multiple frequency bands, though that is future research, and outside the scope of the current paper.

Horn and Opher [12] and Usher et al [13] have investigated models involving competition in a layer of oscillatory units. They used pre-determined weights in their networks and a single inhibitory neuron to demonstrate that there are a limited number of possible sustainable oscillations. Our model is different in that there are multiple inhibitory connections, and these are all learnt through an unsupervised learning scheme.

Bays and Husain [14] examined subjects’ memory of location and orientation information of visual objects. They suggest that visual memory resources are limited, and needs to be shared amongst multiple objects. Hence, increasing the number of objects decreases the precision with which the objects are stored.

The capacity limit that we have investigated in the sensory processing stream could also bear relevance to research in multi-tasking, where similar limits have been found. Liu [15] developed a model using queuing theory to account for human performance in concurrent spatial and visual tasks. He investigated the parallel allocation of multiple resources. The work presented in our paper could dovetail with such a queuing model, such that the visual input consists of a few readily separated streams of objects based on the synchronization between individual processing units.

Recordings of neural activity, such as event-related potentials are being used to shed light on the capacity limits in tracking multiple moving objects. Drew and Vogel [16] have investigated neurophysiological mechanisms that seek to explain why most individuals can track only approximately four objects simultaneously. They observe that the capacity of attentional tracking is approximately four items, and is very similar to the capacity of visual working memory, which is three to four items. They point out that it is likely that similar capacity-limited mechanisms underlie the performance of both memory capacity and tracking capacity. Though the results of our research have been obtained for the simultaneous representation of multiple objects, it is likely to
inform similar capacity related research in other areas such as attention and tracking. This is because our model is not making any specific assumptions about the types of inputs presented. It operates on the principle of unsupervised learning and the ability of synchronized oscillations to convey relatedness information.

Thus our model is general and could be used to model visual memory and attention, though that is not the focus of the current paper. We would expect the same type of capacity limit to be found, as our limit does not depend on the nature of the inputs, but on the ability of a network to separate information about object relatedness in a phase space.

We would like to caution that the results presented in this paper are based on a specific network with a limited number of inputs. Further research and experimentation is required to generalize our result to larger networks and larger input sets.

The results shown in Figure 11 introduce an issue that is related to capacity, which pertains to the accuracy with which simultaneous objects can be represented. Thus, we have to consider not only the capacity of the system, but also the accuracy with which the system can perform at a given capacity. Hence we need to ask the question: how accurately can combinations of two or more objects be simultaneously represented?

Further research in the direction of investigating capacity limits could have broad implications, as the capacity to control attention, maintain information in working memory and multitasking has been linked to higher cognitive functions and intelligence [17], [18].

V. Conclusion

In this paper we have presented a framework for the quantitative modeling of factors affecting capacity limits. We examine the dynamics of an oscillatory network that groups common visual features of an object, such that these features are synchronized with the same phase. We show that multiple objects can be simultaneously presented to the network, and subsequently separated based on their phases. We demonstrate a capacity limit in such a sensory coding network, where up to four simultaneously presented objects can be separated. This result based on quantitative modeling matches results obtained in the field of psychology. Further research is required to generalize our result over larger size networks and larger input sets.

We also show that by narrowing a tuning function that governs the network dynamics, we can achieve higher separation accuracy. However, this comes at the cost of using a higher number of settling iterations, where transients are allowed to die down. Thus, we show that there is a tradeoff between the width of the tuning function and the time scale over which the network operates. Future research in this direction will illuminate the interplay of different factors that produce capacity limits in neural systems. This could have broad implications in understanding attention, working memory, and multitasking, all of which impact higher cognitive functions and intelligence.

Acknowledgement

We are grateful for comments from the anonymous reviewers.

REFERENCES

Fig. 8. This figure shows the distributions of phases in the upper layer. The distributions in Fig. 5(F), Fig. 6(F), and Fig. 7(F) are redrawn after omitting the bin at phase angle zero. The distributions have also been normalized to create an estimate of the probability density function for non-zero phase angles. (A) shows the probability density function of non-zero phase angles when two objects are combined. (B) shows the probability density function of non-zero phase angles when three objects are combined. (C) shows the probability density function of non-zero phase angles when four objects are combined.

Fig. 9. The kurtosis of the probability density function of phase differences is plotted against the number of objects presented to the system.

Fig. 10. Different tuning functions investigated.

(A) Original tuning function, defined by
\[ y = (1 + \cos(x)) \]

(B) Narrower tuning function.
The curve is defined by \[ y = (1 + \cos(x))^3/4 \]
(A) Narrow tuning function, using 200 iterations for settling.

(B) Narrow tuning function, using 400 iterations for settling.

(C) The result in (B) is superposed with the result in Figure 4

Fig. 11. Performance variation with respect to tuning.