COMPARISON OF MOTOR-BASED VERSUS VISUAL SENSORY REPRESENTATIONS IN OBJECT RECOGNITION TASKS

A Thesis

by

NAVENDU MISRA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2005

Major Subject: Computer Science

COMPARISON OF MOTOR-BASED VERSUS VISUAL SENSORY REPRESENTATIONS IN OBJECT RECOGNITION TASKS

A Thesis

by

NAVENDU MISRA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Yoonsuck Choe
Committee Members,	Karen L. Butler-Purry
	Frank Shipman
Head of Department,	Valerie E. Taylor

August 2005

Major Subject: Computer Science

ABSTRACT

Comparison of Motor-based versus Visual Sensory
Representations in Object Recognition Tasks. (August 2005)

Navendu Misra, B.S., The University of Texas at Austin
Chair of Advisory Committee: Dr. Yoonsuck Choe

Various works have demonstrated the usage of action as a critical component in allowing autonomous agents to learn about objects in the environment. The importance of memory becomes evident when these agents try to learn about complex objects. This necessity primarily stems from the fact that simpler agents behave reactively to stimuli in their attempt to learn about the nature of the object. However, complex objects have the property of giving rise to temporally varying sensory data as the agent interacts with the object. Therefore, reactive behavior becomes a hindrance in learning these complex objects, thus, prompting the need for memory.

A straightforward approach to memory, visual memory, is where sensory data is directly represented. Another mechanism is skill-based memory or habit formation. In the latter mechanism the sequence of actions performed for a task is retained. The main hypothesis of this thesis is that since action seems to play an important role in simple perceptual understanding it may also serve as a good memory representation. In order to test this hypothesis a series of comparative tests were carried out to determine the merits of each of these representations. It turns out that skill memory performs significantly better at recognition tasks than visual memory. Furthermore, it was demonstrated in a related experiment that action forms a good intermediate representation of the sensory data. This provides support to theories that propose that various sensory modalities can ideally be represented in terms of action. This thesis successfully extends action to the role of understanding of complex objects.

To my parents Girja P. Misra and Malti Misra

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor, Dr. Yoonsuck Choe, for his immense support and guidance from the initiation of this research till the very end. He has been constantly accessible and has provided numerous approaches for a range of problems that I encountered while conducting this research.

My lab mates, Yingwei Yu and Heejin Lim, were very helpful and provided various ideas during the initial setup of the research. Later, they took time off from their schedules to assist me in the proper presentation of the work performed in this research.

I would also like to thank my other committee members, Dr. Karen Butler-Purry and Dr. Frank Shipman, for their guidance. Dr. Heather Bortfeld was very kind to attend my thesis defense and offer suggestions on improving my thesis.

Finally, I would like to thank my parents for their complete support in everything that I do.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

A. Problem overview 2
1. Role of action in perceptual systems 2
2. Use of memory to enhance perceptual understanding . 3
3. Approach 4
B. Outline of the thesis 6

II BACKGROUND 7

A. Sensory invariance driven action (SIDA) 7
B. Sensorimotor approach to vision 9
C. Rhythm recognition 10

III METHODS 12

A. Input preparation 13
1. Shape generation 13
2. Visual memory – 2D array 14
3. Skill memory – spatio-temporal pattern (STP) 18
B. Neural network experiments 25
1. Backpropagation 25
2. Training and testing 28

IV RESULTS 30

A. Visual memory vs. skill memory in recognition tasks . . . 30
B. Skill memory with variations 32
1. Smoothing window size 32
2. Random starting points 33
3. Trajectory noise 36
C. Action as an intermediate representation 38

V DISCUSSION 42

A. Properties of STP 42
B. Potential limitations 43
C. Relation to memory in humans 47

CHAPTER Page

D. Future work 48
E. Contributions 48

VI CONCLUSION 50

REFERENCES 51

VITA 53

LIST OF FIGURES

FIGURE Page

1 **The representations for circle and triangle.** The gray arrow indicates the starting point of the agent on the 2D object representation for skill memory (left column). From this location the agent will start its movement around the shape and store the sequence of actions performed to navigate the shape. This generates the spatio-temporal pattern (STP) which is shown on the right column of the figure. 5

2 **The difference between the external and internal observer models.** The diagram for external observer, on the left, demonstrates how the input I creates a spike pattern S. Since the observer has access to both of these it can try to come up with a model that explains their correlation. However, in the internal observer model, on the right, the agent has no access to the outside environment and as such can only monitor its internal sensory state or spike pattern. [1] 8

3 **The different components of SIDA agent.** The visual field receives stimulus from a section of the visual environment. This is fed to the filter bank that activates the sensor array (of orientation filters). The agent then performs certain actions based on the sensory state, which may in turn affect the sensory state. [1] 9

4 **The visual representation of circles.** This sequence of figures illustrates the range of variations that are performed on the circle shape in the visual memory (2D array) representation. 15

5 **The visual representation of triangles.** This sequence of figures illustrates the range of variations that are performed on the triangle shape in the visual memory (2D array) representation. . 16

6 **The visual representation of squares.** This sequence of figures illustrates the range of variations that are performed on the square shape in the visual memory (2D array) representation. 17

FIGURE Page

7 **The STP for a circle shape visualized along the diagonal (so as to prevent overlapping of vectors).** (*a*) This plot demonstrates a linear ordering of the spatio-temporal pattern for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors. (*b*) This plot demonstrates a linear ordering of the spatio-temporal pattern for a circle shape after smoothing is applied. Here the action vectors have been smoothed to display a continuous variation in actions. 19

8 **The STP for a circle shape visualized at the coordinate where these actions were performed.** (*a*) This plot demonstrates a 2 dimensional view of the spatio-temporal pattern for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors in the stairstep-like formation. (*b*) This plot demonstrates a 2 dimensional view of the spatio-temporal pattern for a circle shape after smoothing is applied. The smoothed action vectors show a more intuitive sequence of actions. 20

9 **The STP for a triangle shape.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, but for a triangle. . 21

10 **The STP for a triangle shape plotted in 2D.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 8, but for a triangle. 22

11 **The STP for a square shape.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, but for a square. . . 23

12 **The STP for a square shape plotted in 2D.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 8, but for a square. 24

13 **Learning curve for visual and skill memory.** This plot shows the average learning curve for both skill and visual memory (10 trials each). From this plot we can see that skill memory is learned faster and approaches very close to the zero MSE mark early. On the other hand visual memory still has a higher MSE even after 1,000 epochs. 31

FIGURE Page

14 **The average classification rate of visual and skill memory.** This bar chart shows the average classification rate of skill and visual memory on the test set (± 95 % CI). Skill memory has a smaller variance and higher average classification rate representing a more consistently good performance as opposed to visual memory. . 32

15 **The effect of smoothing on the classification rates for skill memory**. This bar chart shows the effect of varying window size on the average classification rate (± 95 % CI). Window size 3 yielded the most optimal performance with the lowest variation. . . . 33

16 **The 2D view of the STP for a square with smoothing.** The two arrows point to the two different locations where the STP sequences may have started. 34

17 **The linear ordering of the spatio-temporal pattern for a square with differing start points.** The plot in (*a*) shows how the STP will appear if the STP generation started from the original start point. The plot in (*b*) shows how the STP will appear if the STP generation started from the new start point (shown in Fig. 16). The dashed arrow in (*b*) points to the original starting location. The only difference with the original version is that the STP is shifted, but that is enough to affect the classification accuracy. 35

18 **The effect of increasing start points on the classification rate for skill memory.** This bar chart shows the change in the average classification rate as the number of start points is increased (± 95 % CI). As the number of random start points is increased, the average classification rate steadily drops, but clearly skill memory performs better than visual memory. 36

19 **The STP with noise for a square shape**. The plot shows a 1 dimensional view of the spatio-temporal pattern for a square after the application of random noise (noise factor 0.1) and after smoothing. 37

20 **The STP with noise for a square shape**. The plot shows a 2 dimensional view of the spatio-temporal pattern for a square after the application of random noise (noise factor 0.1) and after smoothing. Notice how the shape traced out is still quite close to the original square. 38

FIGURE Page

21 **The effect of noise on classification rate.** This bar chart shows the effect of increasing noise on the average classification rate of skill memory (± 95 % CI). Visual memory is shown here as a baseline. The detrimental effect of noise can clearly be observed from this bar chart. However, notice that skill memory is quite resilient to noise and outperforms visual memory at noise factor of 0.5. 39

22 **The learning curve for the different mappings.** This plot shows the learning curve for each mapping. From the learning curves we can infer how well each mapping performs with respect to other mappings. As expected the visual to action mapping performs well. The visual to action mapping is also relatively good. . 41

23 **The plot of PCA projection for visual and skill memory.** (*a*) PCA for 2D array representation (visual memory) of the input data along the first two principal axes is shown. (*b*) PCA for STP representation (skill memory) of the input data along the first two principal axes is shown. 44

24 **The projection on the two principal axes for skill memory with varying noise factor.** (*a*) PCA for skill memory with a lower noise factor of 0.1 along the first two principal axes is shown. (*b*) PCA for skill memory with a high noise factor of 0.5 along the first two principal axes is shown. 45

25 **The projection on the two principal axes for skill memory with varying number of random start points.** (*a*) PCA for skill memory with two random start points for STP generation along the first two principal axes is shown. (*b*) PCA for skill memory with four random start points for STP generation along the first two principal axes is shown. 46

CHAPTER I

INTRODUCTION

Earlier study on Sensory Invariance Driven Action (SIDA) [1] has demonstrated the importance of actions for an agent trying to learn about the environment when it is only able to access its internal sensory state. The internal sensory state is a representation of the external environment that arises from the application of sensory filters that extract meaningful properties from the environment. The SIDA agent is designed to carry out actions having particular patterns that maintain invariance in its internal sensory state. For the basic SIDA agent this invariance maintenance required a direct comparison of the current internal state with the immediate past sensory state. This immediate comparison allowed the SIDA agent to learn about simple stimulus properties, where there are no variations in sensory property along an object. However, this reactive behavior presented a hindrance in understanding more complex objects, which may contain variations in the sensory state. How can we address this problem? One possibility is to allow SIDA to use some form of memory.

There are several different forms of memory. The most straight-forward memory is visual memory. Visual memory is a direct storage of the raw sensory state, thus visual memory is based only on sensory representations. A different kind of memory system, skill memory, stores the action sequence that the agent performs while investigating the object. Unlike visual memory, skill memory is based on motor representation. The main hypothesis of this thesis is that since action seems to serve a fundamental role in allowing the agent to learn about simple properties in its environment, it may also be a good medium for representing and understanding more

The journal model is *IEEE Transactions on Neural Networks.*

complex object properties. That is, action may serve as a good basis for memory.

Memory system is a critical component in the building of autonomous agents capable of complex behavior. One of the reasons for this is that studies of animals have demonstrated that the development of memory has been crucial [2]. It has also been demonstrated that animals routinely need the capacity for spatial navigation and context-dependent learning that is provided by memory [3]. These are some of the intelligent behaviors that one might expect from autonomous agents and thus the study of memory (in general) and building systems that accomplish it is important. The question that this all leads to is, is one of these forms of memory more suitable for representing and *understanding* stimulus properties?

A. Problem overview

For a better understanding of the core problem addressed in this thesis, a brief review is provided of the current studies that put an emphasis on action. Then their limitations with regards to memory are detailed, followed by an introduction to the approach taken in this thesis.

1. Role of action in perceptual systems

Action has been an essential element in the development of learning in agents. This was demonstrated in the example of SIDA [1]. In SIDA it was shown how perception and action can be bound together to provide a framework for learning while using a simple yet powerful concept of sensory invariance. This important concept of linking perception with action through invariance was also demonstrated in the experiments conducted by Philiopona et al. [4] (although in their work invariance was used in a different context) which built upon the theoretical framework of [5].

Recently, a system for rhythm recognition was also developed by Buisson [6]. This system worked by producing anticipated notes that would be matched to what was observed in the environment. Essentially, the system did the note matching by having a population of internal rhythms that it maintained. The core contribution of this work was that it showed how even complex sequence of notes can be recognized by a relatively simple note-matching system. The generation of these notes can be considered as an action. This is therefore another example of where action can be a useful tool in learning. The approaches mentioned above will be covered in more detail in the background chapter.

This thesis aims to further investigate the role of action in perceptual understanding, when more complex action sequences can be retained (as in habit formation). This work is expected to show that skill-based memory developed from an application of SIDA can have several beneficial properties compared to those based on direct sensory representations.

2. Use of memory to enhance perceptual understanding

SIDA performs actions that maintain invariance in its internal sensory state. This is achieved by performing a direct comparison of its immediate past and present internal sensory states. For a simple object (e.g., a straight line), this simplistic comparison of current and past sensory states suffices. However, for a complex object the very short term comparison may not be enough. A square is an example of such a complex object. For a square, the agent needs to remember that it has seen lines in varying orientations (horizontal and vertical) as the agent's visual field traverses different parts of the object. This is quite different from an environment where there is only a straight line and when only the same sensory state is activated no matter which part of the line is in the agent's visual field. The current version of SIDA has no capacity for remembering these variations of sensory state over a longer stretch of time. As a consequence, improving the SIDA agent to learn about complex objects may require allowing it to remember a series of past internal states. In other words, it needs a memory system.

3. Approach

This thesis investigates two types of memory representations that can be employed by SIDA; sensory-based visual memory and motor-based skill memory. Visual memory stores raw sensory states that are caused by immediate sensory activation which captures the entire visual field. This is similar to the pattern of activity in the retina in the eye. This form of memory stores the direct sensory mapping into the memory space. A quick examination will yield that the memory space will look exactly like the image that the agent had on its visual field, as demonstrated in the left column of Fig. 1. Skill memory on the other hand, is the storage of action sequences that the agent performed while interacting with the object through the invariance maintenance criterion mentioned earlier. (The action sequence is hereafter referred to as spatio-temporal pattern (STP).) The right column in Fig. 1 shows the STP for the corresponding shape on the left.

For this thesis memory representations using skill and visual memory for a number of objects will be evaluated. A comparison will be made of how each of these perform on recognition and mapping tasks using a feedforward neural network using gradient descent backpropagation. The performance difference between the two memories will demonstrate whether more complex meanings can be learned more easily when it was based on action-based representations.

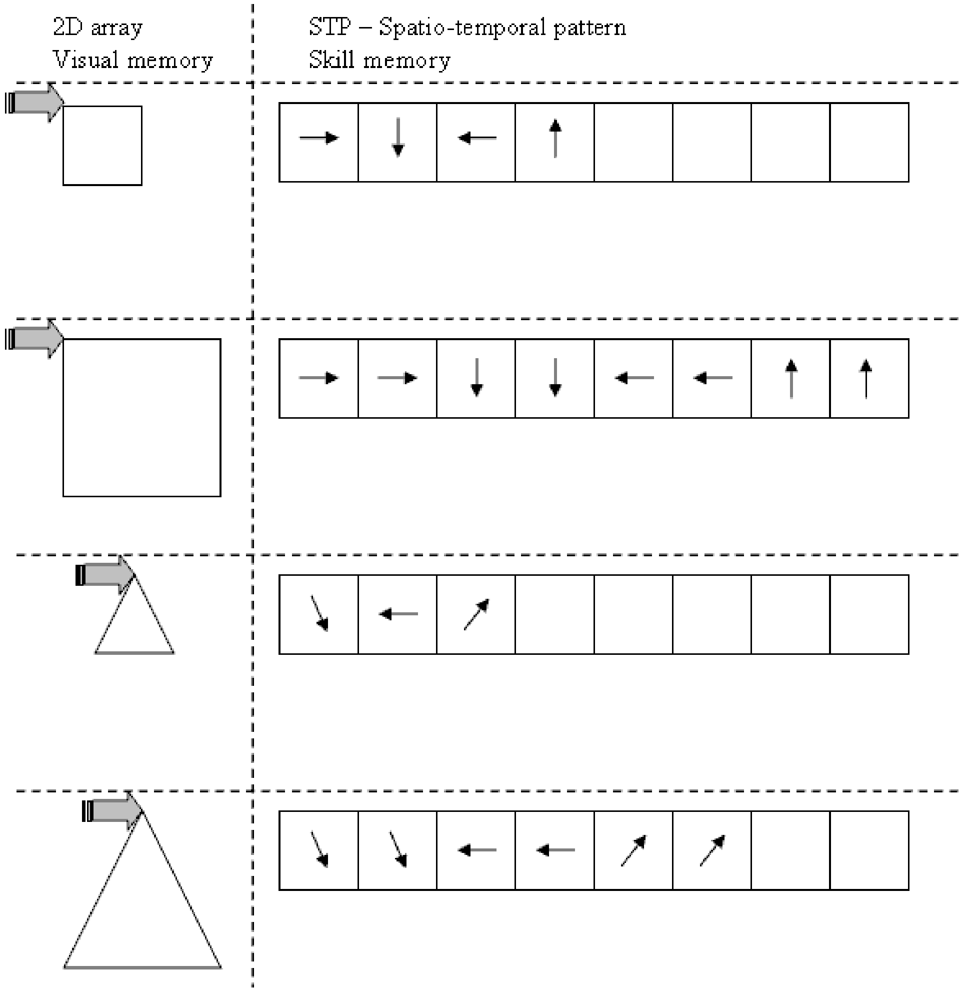


Fig. 1. **The representations for circle and triangle.** The gray arrow indicates the starting point of the agent on the 2D object representation for skill memory (left column). From this location the agent will start its movement around the shape and store the sequence of actions performed to navigate the shape. This generates the spatio-temporal pattern (STP) which is shown on the right column of the figure.

B. Outline of the thesis

Chapter II of the thesis provides the necessary background for comprehending the importance of action in perceptual understanding. Chapter III deals with the actual setup of the experiments for the test of the core differences between skill and visual memory. Chapter IV will detail the results of the experiments performed. The following Chapter V, discussion, will provide interpretations and analyses of the results. Finally, Chapter VI concludes the thesis with a brief outlook.

CHAPTER II

BACKGROUND

In this chapter, methods that have used action as a fundamental component in perceptual/conceptual understanding will be reviewed. An influential paper by Aloimonos et al. [7] claimed that vision does not exist independent of action, i.e., vision and action may be intricately linked. This view leads to the establishment of frameworks such as SIDA, where action and sensory invariance lead to the formation of a learning system. In fact Karl Lashley, a prominent psychologist, had suggested, that vision cannot be explained without using postural-kinaesthetic considerations and that vision is in fact linked to action and thus the correlation between action and perception should be studied [8].

A. Sensory invariance driven action (SIDA)

SIDA was developed primarily to answer the question of how to associate meaning to a spike of a sensory neuron without direct access to the environmental stimulus. A spike is an electrochemical pulse that is generated upon the activation of a neuron. In the case of the sensory neuron the spike is caused by the stimulus received by the sensory neuron. Most of the other methods that tried to associate meaning to spiking neuron had tried to come up with intricate correlations between the raw sensory stimulus that caused the spike pattern and the spike pattern itself. This is illustrated in Fig. 2. The figure on the left is the traditional approach where the observer has access to both the input I and the spike pattern S. In this approach the observer can find a correlation between the input and the spike pattern. However, an important question at this point is, how can the brain do this? Let us take the example of a brain composed of just a single neuron. This brain does not have direct

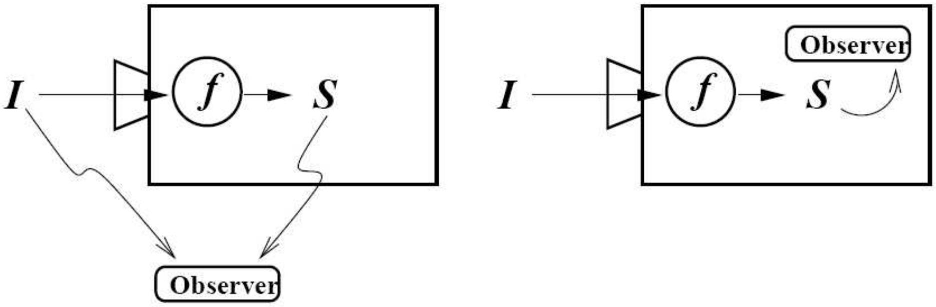


Fig. 2. **The difference between the external and internal observer models.** The diagram for external observer, on the left, demonstrates how the input I creates a spike pattern S. Since the observer has access to both of these it can try to come up with a model that explains their correlation. However, in the internal observer model, on the right, the agent has no access to the outside environment and as such can only monitor its internal sensory state or spike pattern. [1]

access to the input. The right part of Fig. 2 shows this. From this we can see that the traditional approach fails to explain how the brain may actually be associating meaning to the spike pattern that it receives.

SIDA overcomes this problem by using action and the principle of sensory invariance to allow an agent to learn stimulus properties of the environment by just monitoring its internal sensory state. This allowed the internal observer to correlate the actions that it performed with the observed spike pattern. As a result, SIDA grounds its internal sensory state on its actions.

The core features of SIDA are illustrated in Fig. 3. Here the agent sees a line at a particular angle that activates the corresponding orientation filters which is tied to a unique sensory state. The agent then performs actions by the usage of the invariance criterion mentioned above. SIDA agent was designed to initially start off by trying a large variation of actions. After a while, it will determine that a particular action pattern maintains invariance in its internal state. From this, it will learn to associate

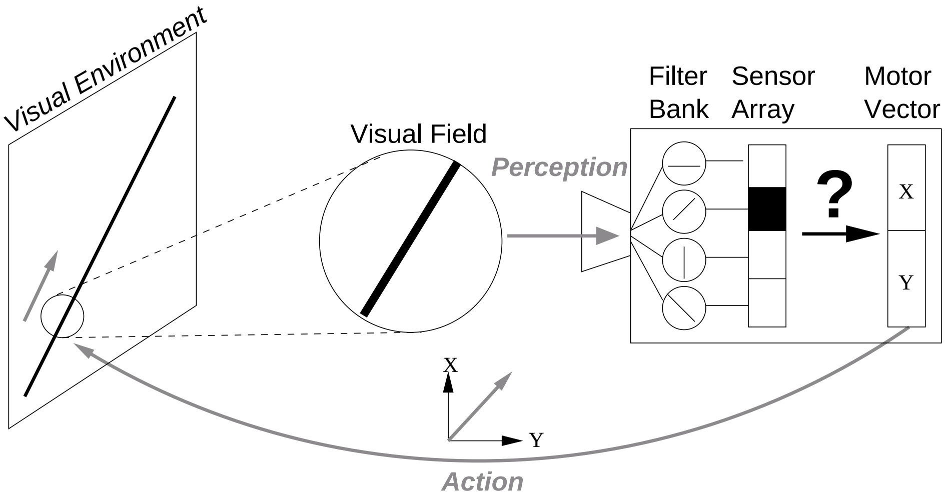


Fig. 3. **The different components of SIDA agent.** The visual field receives stimulus from a section of the visual environment. This is fed to the filter bank that activates the sensor array (of orientation filters). The agent then performs certain actions based on the sensory state, which may in turn affect the sensory state. [1]

particular actions with specific sensory responses and thus infer properties of the environment. Invariance is maintained by comparing the current sensory activation with the immediate past sensory state. The agent will then perform an action that will lead to the successful matching of the two, where the property of the resulting action reflects that of the stimulus. In sum, SIDA allows an internal observer to learn about external stimulus properties through action, when only internal state information is available.

B. Sensorimotor approach to vision

In the same line as SIDA, another work, the sensorimotor approach, has demonstrated the importance of action in vision [5]. Traditional views on vision suggested that vision consisted of creating an internal representation of the external world and the challenge therefore was the creation of an appropriate mapping of the external world with the internal representation. The research of O'Regan and Noë suggests that perception can only be understood in terms of action. The authors have proposed that consciousness or understanding of objects in an environment comes about as series of actions are performed as opposed to earlier methods that viewed this as a simple mapping from sensory input to motor output.

Philipona et al. [4] take this idea further by demonstrating how the brain can come to understand its environment if it does not have any a priori information about its sensors. The paper demonstrates that the brain is able to conclude that there is a distinction between the aspects that it can control, its body, and features that it has no direct control over, its environment. Essentially, there is a class of compensable actions. These compensable actions can compensate for changes in the environment and as a result keep the sensory information constant. The core subject of the paper is the determination of the dimensionality of space that an organism is situated in. The main point is that by using action only (such as compensable actions) the agent can determine these dimensions, thereby allowing the brain to understand the environment by only performing actions and relating the changes in the sensory states.

C. Rhythm recognition

Perception and action was again important in the research by Buisson [6]. He demonstrated how the simple act of generating notes leads to the formation of a model that can dynamically and closely recognize complex environmental input. Buisson argues that the mental representation cannot simply be copies of the world, which as indicated in the paper was initially pointed out by Piaget. (Note that Philipona et al. also had similar views.) One of the core differences between rhythm recognition and the more traditional methods is that it is able to capture the core temporal qualities of the environmental input, by the clever usage of its population of rhythm generators, where sequence of notes are learned by attunement to the temporal pattern present in the sound.

This theory is demonstrated by a Java-based program. The program does its pattern matching in quite a straight forward way using what the paper calls the sensory motor scheme (SMS). An SMS is just a sequence of notes that the system is internally playing and determining if it is tuned to the observed environmental stimulus. Initially, the system starts off with a default SMS. If it matches the observed state of the environment, the particular SMS is deemed to be successful. Then more SMSs with similar action sequences are produced. The production of an SMS is controlled by a mutation factor that alters with the success of the particular SMS. If the SMS fails then its mutation rate is increased and more deviations in its action pattern are accepted. This dynamic system generates multiple SMSs running in parallel at any given time.

In this experiment the internally playing notes can be thought of as internal, virtual actions and the comparison of the action (notes) to the actual rhythm the sensory feedback. In fact, its rhythm matching scheme not only uses action but the sequence of actions actually forms a type of habit which allows the system to predict future states of the environment. This form of action-based recognition was a motivation toward the skill memory used in this thesis.

CHAPTER III

METHODS

In order to compare the relative merits of the two memory representations, a set of systematic tests was carried out. Objects (circles, square, and triangles) were represented in each representation, STP (for skill-based memory) and 2D array of pixels (for visual memory), of varying sizes and locations in an equal dimensional space. One problem associated with skill-based representation is that the objects are represented with varying length of STP depending on the size of the object. However, the dimensionality that both representations (skill and visual) occupy should be the same so as to allow for a fair comparison between the two. In order to overcome this problem, resizing has to be done on each STP. Another issue with STP was that it had fixed range of actions that produced discrete stairstep-like aliasing in the sequence of action vectors. In order to overcome this problem, smoothing was applied on the STP. This resulted in the neighboring action vectors having less dramatic changes in their angles. In another attempt to have fair comparison between visual and skill memory, the 2D array representing the visual memory was also made smoother (i.e., blurred).

Once the representations had been constructed for the various objects, the data was partitioned into training and test sets. The training set was provided to a feed-forward neural network trained by backpropagation and the performance measured with regards to mean square error (MSE) as well as the time elapsed in the training of the neural network until an asymptotic error level was reached. Then the test set was presented and the degree of generalization measured as the average classification rate.

In order to rigorously test the differences between the two memories, certain variations were considered that had the potential of impacting the relative performance of each of these memories. One of these variations was the random start points for STP generation in skill memory. Without random start points what happens is that the action sequences follow the same order for the same shapes. Random start points allow for the action sequences to be out of order. This may be analogous to how visual memory performs on translated objects. Another interesting variation came in the form of noise in the trajectory. This had the effect of randomly perturbing action vectors in a particular action sequence with a certain factor.

To test whether action can serve as a good intermediate representation, i.e. whether action can serve as a canonical representational scheme when there are varying types of sensory input to deal with; vision, audition, touch, etc., I tested the relative ease of mapping between two different kinds of representations (visual and action-based). This involved testing the mapping from visual to action and action to visual. This was done to test our theory that action may be a good intermediate representation for sensory data.

The following sections provide details about this setup.

A. Input preparation

The input to the neural network was prepared by randomly generating the STP and the 2D array representation for the three different shapes. It was ensured that each of the representations was defined to have exactly the same number of dimensions.

1. Shape generation

The generation of the three types of shape was done by following a simple algorithm. These algorithms were constructed using a LOGO-like language [9]. In this language instructions for navigating a space is provided by a fixed range of actions, i.e., 'turn left', 'move forward', and so on that are then plotted. This language construct was adapted for the formation of the shape generation algorithms in the following manner. Initially a starting point was chosen for each of the figures. Then a series of steps were produced to allow for a full rendition of the entire shape. The algorithm was parameterized so as to produce images that were scaled and translated. As different coordinates were traversed the corresponding 2D array points were marked. The benefit of the LOGO-like algorithm was that it allowed for the easy capture of the STP for the particular shape. This was the case because the actions produced to traverse the object could be captured as the direct representative STP for a figure. Since each of the algorithms was parameterized, a sequence of random values for scaling and translating the images were provided.

2. Visual memory – 2D array

Visual memory representation is a direct copy of the sensory data. As a result, when the 2D array representation for the figure is visualized it appears like the figure itself. All the shapes for the 2D arrays were generated using the algorithm described above. The output was a two dimensional array with the pixels of the edge of the figure set to be one and the rest of the figure to zeros. On this a Gaussian filter was applied. This caused the values in the array to have more continuous values. The primary reason for this last step was to have a more fair comparison between 2D array and STP. The normalized range of values was between 0 and 1 and the resultant size of the array was 30 × 30.

Fig. 4, Fig. 5, and Fig. 6 provide examples of the different shapes in the 2D array representation. Note that there is a less sharp transition between the edge of the figure and the background. This is achieved from the smoothing effect of the

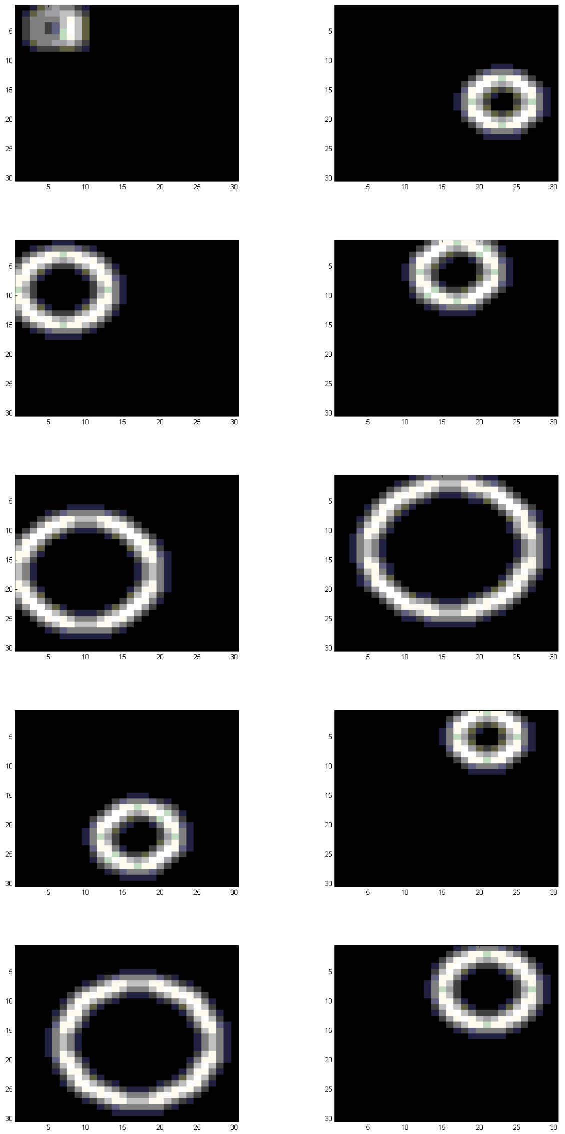


Fig. 4. **The visual representation of circles.** This sequence of figures illustrates the range of variations that are performed on the circle shape in the visual memory (2D array) representation.

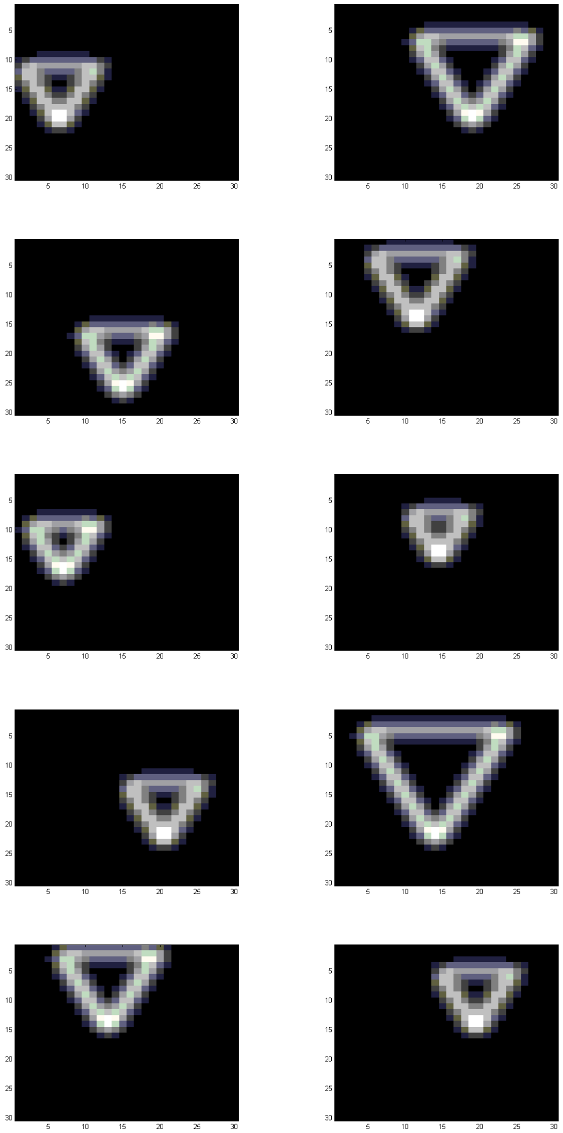


Fig. 5. **The visual representation of triangles.** This sequence of figures illustrates the range of variations that are performed on the triangle shape in the visual memory (2D array) representation.

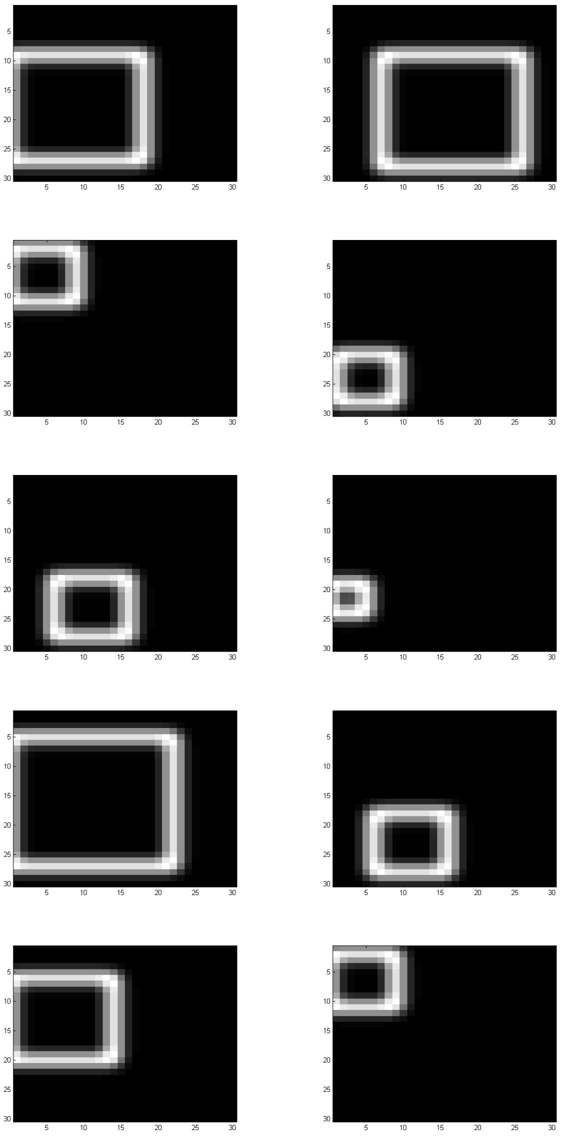


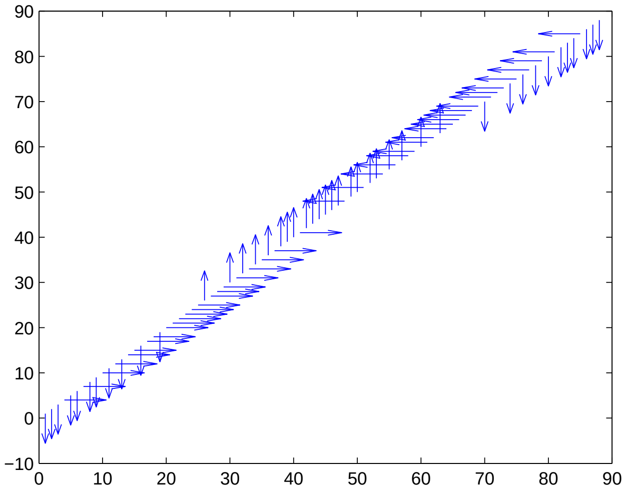
Fig. 6. **The visual representation of squares.** This sequence of figures illustrates the range of variations that are performed on the square shape in the visual memory (2D array) representation.

Gaussian filter mentioned above. Such a blurring enables better generalization.

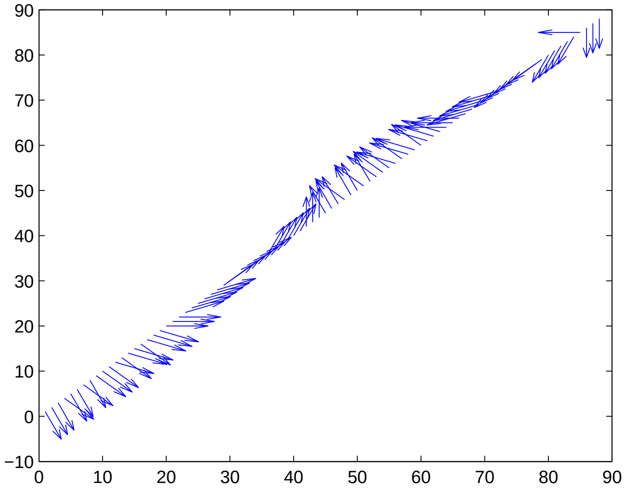
3. Skill memory – spatio-temporal pattern (STP)

Skill memory representation, STP, involves the retention of actions that an agent may perform while navigating the environment. As explained earlier the action sequence was generated by utilizing an algorithm based on the LOGO language. (Note that a similar action sequence is expected when SIDA is used to traverse the object.) The produced output action sequence had four actions; motion north, motion south, motion east, and motion west. This was represented by values 0, 90, 180, and 270 degrees. The values were subsequently normalized to lie between the range 0 to 1. Before normalization the action vectors were smoothed. Smoothing was accomplished by taking an average of values representing the neighboring action vectors, as specified by the size of the smoothing window. This resulted in a less discrete change of action vectors. This difference is illustrated in Fig. 7. In (*a*) the STP is not smoothed and the sequence of actions constructed has stairstep-like variations. However, in (*b*) the actions get averaged to form a new action that is formed by averaging the neighboring action vectors. Fig. 8 shows the action vectors at the coordinates where these actions were performed. Figs. 9 - 12 show the same smoothing effect but for triangles and squares.

Another issue with the STP was that its length was not equal to the fixed dimension size, unlike the visual representation. This meant that resizing of the STP had to be done to match the input dimension ($30 \times 30 = 900$). A very simplistic algorithm was devised to resize or stretch the STP. The end result was that each STP size was of 900 dimensions.

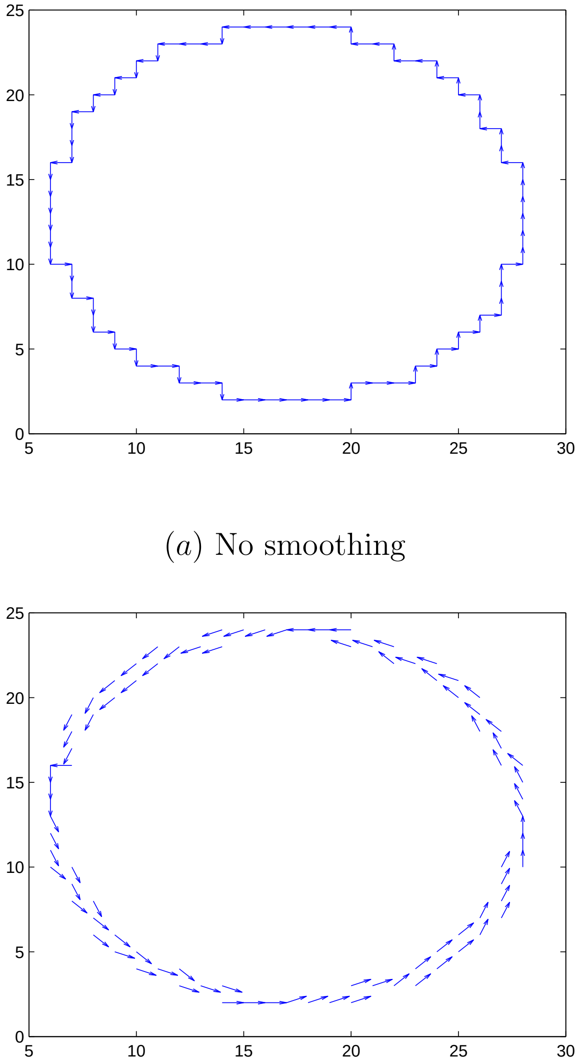


(*a*) No smoothing



(*b*) Smoothing applied

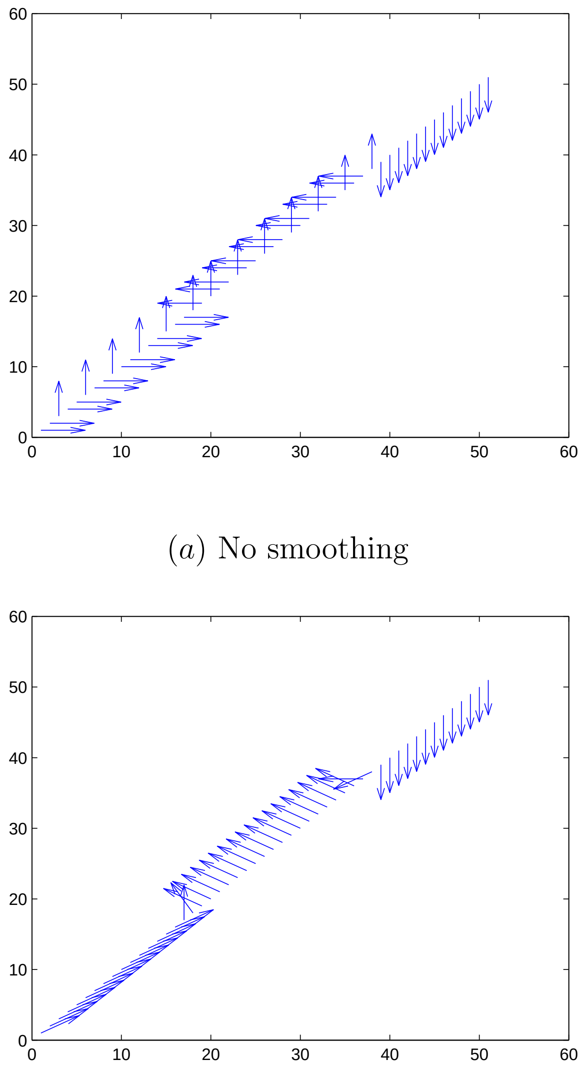
Fig. 7. **The STP for a circle shape visualized along the diagonal (so as to prevent overlapping of vectors).** (*a*) This plot demonstrates a linear ordering of the spatio-temporal pattern for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors. (*b*) This plot demonstrates a linear ordering of the spatio-temporal pattern for a circle shape after smoothing is applied. Here the action vectors have been smoothed to display a continuous variation in actions.



(a) No smoothing

(b) Smoothing applied

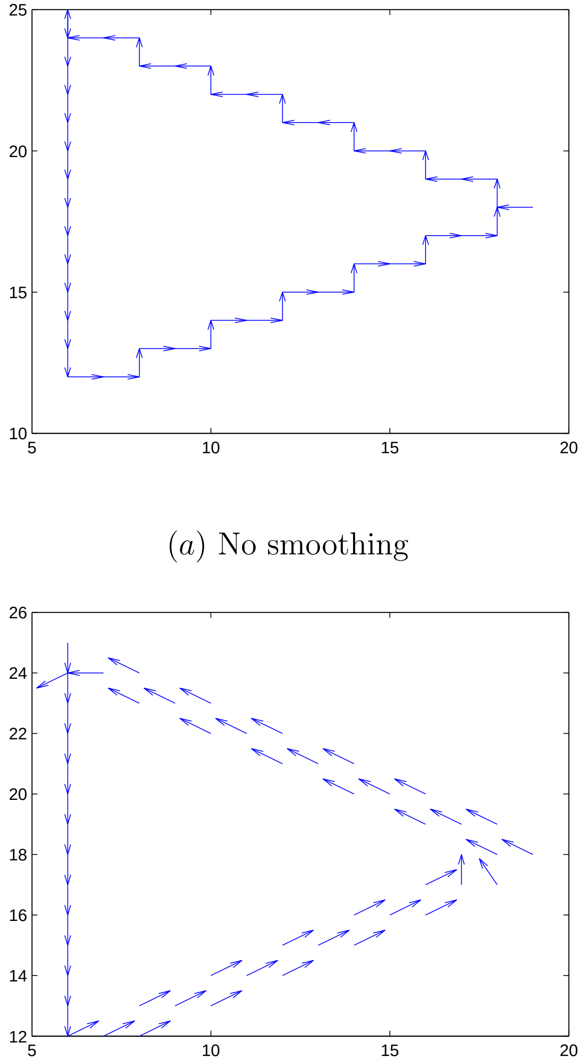
Fig. 8. **The STP for a circle shape visualized at the coordinate where these actions were performed.** (a) This plot demonstrates a 2 dimensional view of the spatio-temporal pattern for a circle shape before any smoothing is applied. Note in this figure the discrete change in action vectors in the stairstep-like formation. (b) This plot demonstrates a 2 dimensional view of the spatio-temporal pattern for a circle shape after smoothing is applied. The smoothed action vectors show a more intuitive sequence of actions.



(*a*) No smoothing

(*b*) Smoothing applied

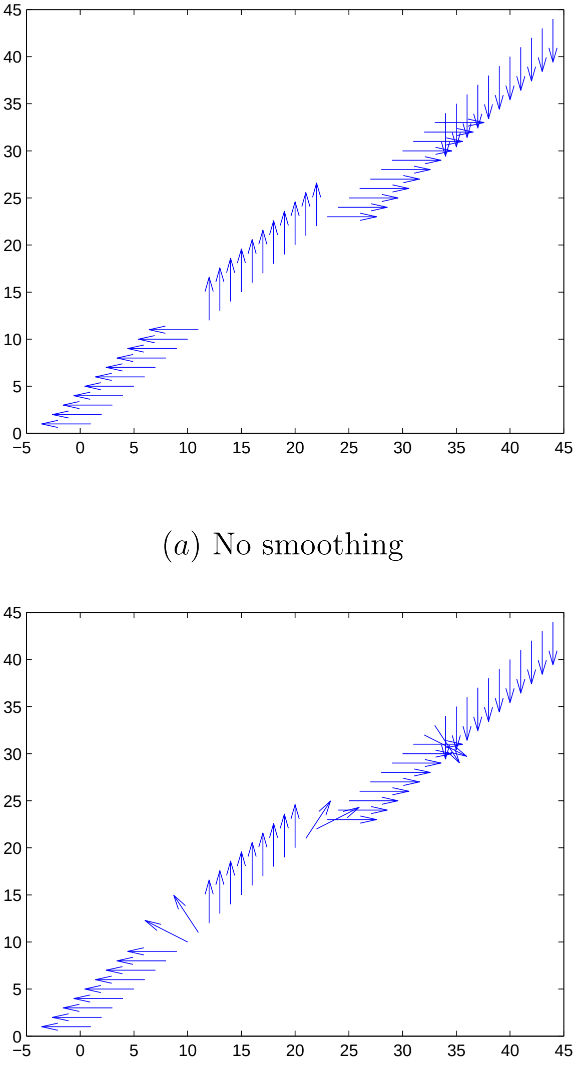
Fig. 9. **The STP for a triangle shape.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, but for a triangle.



(*a*) No smoothing

(*b*) Smoothing applied

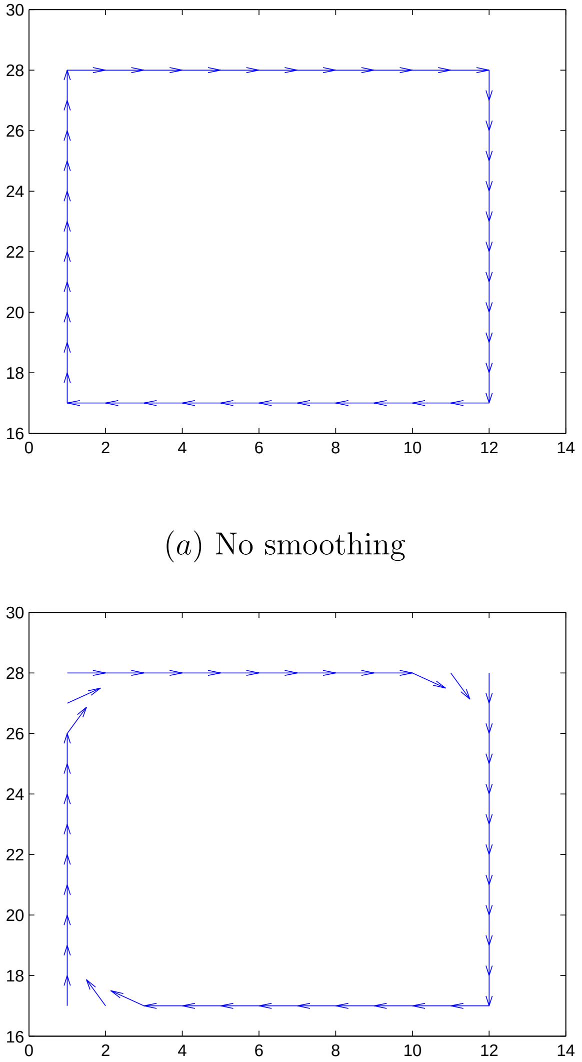
Fig. 10. **The STP for a triangle shape plotted in 2D.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 8, but for a triangle.



(*a*) No smoothing

(*b*) Smoothing applied

Fig. 11. **The STP for a square shape.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 7, but for a square.



(a) No smoothing

(b) Smoothing applied

Fig. 12. **The STP for a square shape plotted in 2D.** These plots demonstrate the same effect of smoothing as mentioned in Fig. 8, but for a square.

B. Neural network experiments

The experiments were formulated by creating one thousand randomly scaled and randomly translated figures for each shape (triangle, circle, and square) with their corresponding STP and 2D array representations. The patterns generated for the STP and 2D array representations were stored in their respective data sets. A portion of this data set (75 %) was used for training a neural network and the rest was used for testing (25 %). Using the testing dataset, the performance of each of the representations was recorded. A detailed explanation is given in the following sections.

1. Backpropagation

This thesis required a general purpose learning algorithm. One such algorithm is the backpropagation algorithm for training artificial neural networks. These artificial neural networks are generally made up of several neurons and each neuron can have weighted connections with other neurons. Each neuron can receive activation (information) from other neurons that it is connected to. These neural networks are also arranged in layers. There are primarily three types of layers; the input layer, the output layer, and the hidden layer. The input layer is activated directly by the sensory data. This activation is then fed to the hidden layer, which does its processing. Then finally the output layer receives the activation, this is where the output of the neural network is noticed. When the artificial neural network learns a particular input-output sequence, the connection weights between neurons are altered to produce the desired output from the output layer. The backpropagation algorithm works by enabling the weights to change in an efficient way for a given sequence of input-output pairs.

The backpropagation algorithm works by performing a step-by-step updating of the weights. In order to do this the algorithm goes through the training data by following the steps below [10]:

1. Initialize the network: The algorithm starts with a default configuration. Then all the synaptic weights between the neurons and threshold levels of the network are set to small randomly distributed numbers.

2. Provide the training examples: Provide the network with an epoch of training examples. A sequence of forward and backward computations specified in steps 3 and 4 is repeated, until performance measures such as MSE level off.

3. Forward computation: The training example will be made of the input $x(n)$ and the desired output $d(n)$, where n is the sequence number in a particular epoch. The activation will be calculated by proceeding forward through the network, layer by layer. The net internal activity $v_j^{(l)}\left(n\right)$ for neuron j in layer l is calculated as:

$$v_j^{(l)}\left(n\right) = \sum_{i=0}^{p} w_{ji}^{(l)}\left(n\right) y_i^{(l-1)}\left(n\right), \tag{3.1}$$

   where $y_i^{(l-1)}\left(n\right)$, refers to the function signal of the ith neuron in the previous layer $l-1$ in the nth iteration, and $w_{ji}^{(l)}$ the synaptic weight to the neuron j in layer l from neuron i in layer $l-1$ and p refers to the number of neurons in layer $l-1$. The function $y_i^{(l)}$ can be sigmoidal or tanh based [11]:

$$y_j^{(l)}\left(n\right) = \frac{1}{1+\exp\left(-v_j^{(l)}\left(n\right)\right)}, \text{or} \tag{3.2}$$

$$y_j^{(l)}\left(n\right) = \tanh\left(v_j^{(l)}\left(n\right)/2\right). \tag{3.3}$$

If the jth neuron is in the first layer then its activation is the same as in the input, where $x_j(n)$ is the corresponding jth element of the input; and same as the output if it is the output layer, where $o_j(n)$ is the value at the corresponding output neuron at the top layer L:

$$y_j^{(0)}(n) = x_j(n), \tag{3.4}$$

$$y_j^{(L)}(n) = o_j(n). \tag{3.5}$$

The error signal $e_j(n)$ is computed by

$$e_j(n) = d_j(n) - o_j(n), \tag{3.6}$$

where $d_j(n)$ is the jth element of the desired response vector.

4. Backward computation: For this, the local gradients (δ) are computed for each j (this is for the sigmoidal function).

$$\delta_j^{(L)}(n) = e_j^{(L)}(n)\, o_j(n)\left[1 - o_j(n)\right], \text{and} \tag{3.7}$$

$$\delta_j^{(l)}(n) = y_i^{(l)}(n)\left[1 - y_i^{(l)}(n)\right]\sum_k \delta_k^{(l+1)}(n)\, w_{kj}^{(l+1)}(n). \tag{3.8}$$

The synaptic weights of the network in layer l are updated according to the generalized delta rule:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha\left[w_{ji}^{(l)}(n) - w_{ji}^{(l)}(n-1)\right] + \eta\delta_j^{(l)}(n)\, y_i^{(l-1)}(n), \tag{3.9}$$

where η is the learning-rate parameter and α is the momentum constant.

2. Training and testing

Matlab's Neural Network toolbox was used to run the backpropagation simulation on the dataset for both visual and skill memory. There were a total of ten runs for each experiment. For each of these trials the training set and the test set were chosen at random from the dataset for each class (circle, square, and triangle). This was composed of a thousand points for STP and 2D array for each of the three figures, resulting in STP and 2D array dataset of three thousand each. For each run, at random three quarters of the dataset was chosen as training data and the rest was used as the test data. This was provided to the neural network, and trained using backpropagation. The neural network had 900 input neurons, 10 hidden neurons in the single hidden layer and 3 output neurons corresponding to the 3 classes. For the last experiment the target vectors were modified to be the actual visual and skill memory representations. In this case the number of output neurons was increased to 900.

Average classification rate was a measure that was used to gauge the relative performance for both visual and skill memory for the test set. The classification rate for each trial recorded the average number of times the actual output deviated from the target vector of the three output neurons. A threshold of 0.5 was set so that if the deviation of the output neuron activation was within this value then the particular input was claimed to be properly classified. The average classification rate was then acquired by running the experiments ten times and taking the mean and standard deviation of the values. Student's t-test was used to measure the significance of the differences.

Another measure is the mean square error (MSE). This value represents the average of all the squared deviations of the output values from the exact target values. MSE gives a general idea of how well the mapping was learned in case hard classification is not possible. The other measure, number of epochs to reach an asymptotic MSE, was gathered to determine the speed of learning for each of the experiments.

CHAPTER IV

RESULTS

In order to evaluate the effectiveness of each of the memory representations there needs to be a comprehensive evaluation of each of the memory systems with respect to the performance measures specified in the previous chapter. These performance measures were used to primarily demonstrate the relative difference between the two memory systems rather than provide a mechanism for absolute comparison with a general pattern recognition approach that may seek to maximize the performance.

A. Visual memory vs. skill memory in recognition tasks

The overall speed of learning, measured using MSE, is illustrated in Fig. 13. MSE values for each of the curves were calculated by taking an average of ten trials for each of the two memory representations. The neural network was allowed to train for one thousand epochs. As can be seen from Fig. 13, the error rate for skill memory is consistently lower than that of skill memory. Also after about 200 epochs the MSE comes close to zero for skill memory while visual memory can only reach an MSE value of about 0.1 after the full period of one thousand epochs. The results clearly demonstrate that the neural network can more easily learn the various STP sequences in skill memory.

The differences between skill memory and visual memory are further emphasized in Fig. 14. Here the average classification rate on the test sets is shown using the bar chart with the error bars representing the 95 % confidence interval. The average classification rate for visual memory was 0.28 while for skill memory it was almost four times higher at close to 0.97. These differences were significant under t-test (p = 0, n = 10). In sum, the action-based skill memory was significantly easier to learn

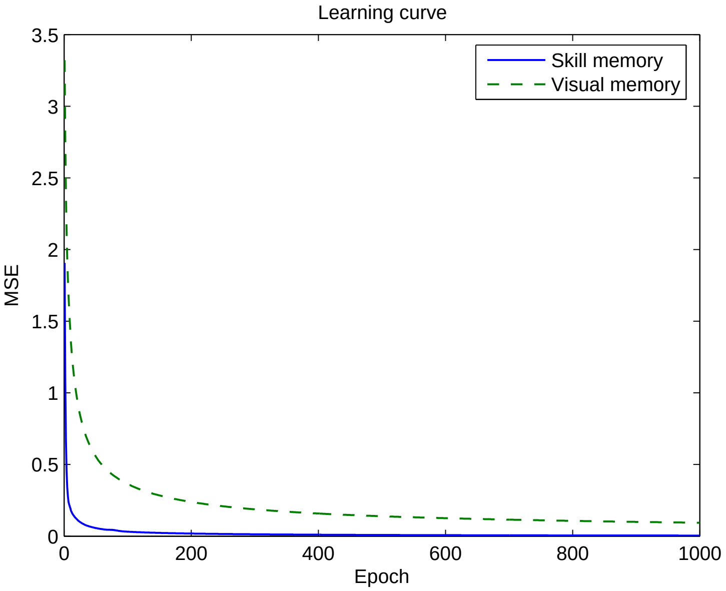


Fig. 13. **Learning curve for visual and skill memory.** This plot shows the average learning curve for both skill and visual memory (10 trials each). From this plot we can see that skill memory is learned faster and approaches very close to the zero MSE mark early. On the other hand visual memory still has a higher MSE even after 1,000 epochs.

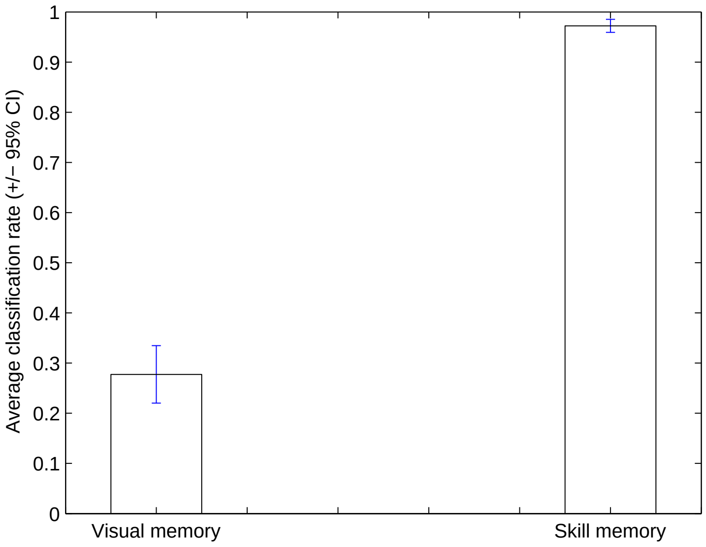


Fig. 14. **The average classification rate of visual and skill memory.** This bar chart shows the average classification rate of skill and visual memory on the test set (± 95 % CI). Skill memory has a smaller variance and higher average classification rate representing a more consistently good performance as opposed to visual memory.

than the visual memory, both in terms of speed and accuracy.

B. Skill memory with variations

The performance of skill memory was measured under variations in the formation of the STP for skill memory. These variations included; (1) changes in the smoothing window size, (2) variations to the number of starting points for a particular STP, and (3) noise in trajectory.

1. Smoothing window size

Smoothing was applied to all the STP that was generated. The effect of this has already been illustrated in the previous chapter. Smoothing has an effect of reducing

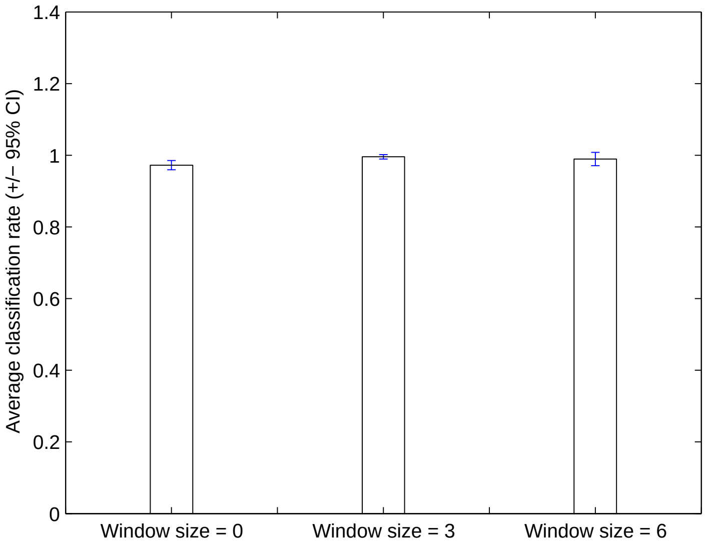


Fig. 15. **The effect of smoothing on the classification rates for skill memory**. This bar chart shows the effect of varying window size on the average classification rate (± 95 % CI). Window size 3 yielded the most optimal performance with the lowest variation.

the average classification rate at a lower threshold, i.e. when the deviation from the desired target is low. However, as the threshold is increased to 0.5, the average classification rate increases slightly with an increase in the window size. However, a further increase in widow size causes the average classification rate to decrease slightly. As a result the default smoothing window size was chosen to be three. Fig. 15 shows how smoothing affects the average classification rate for skill memory. These values were averages taken by performing ten trials. All the differences were significant under t-test ($p < 0.04$, n = 10).

2. Random starting points

Another variation was to test how the classification rate was impacted by starting points chosen for the STP generation. This was implemented by varying the number

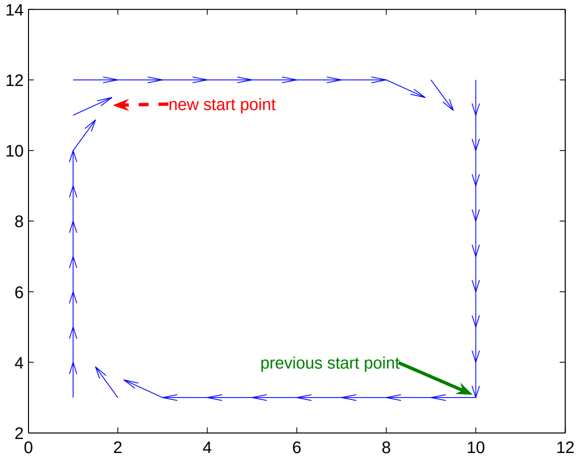
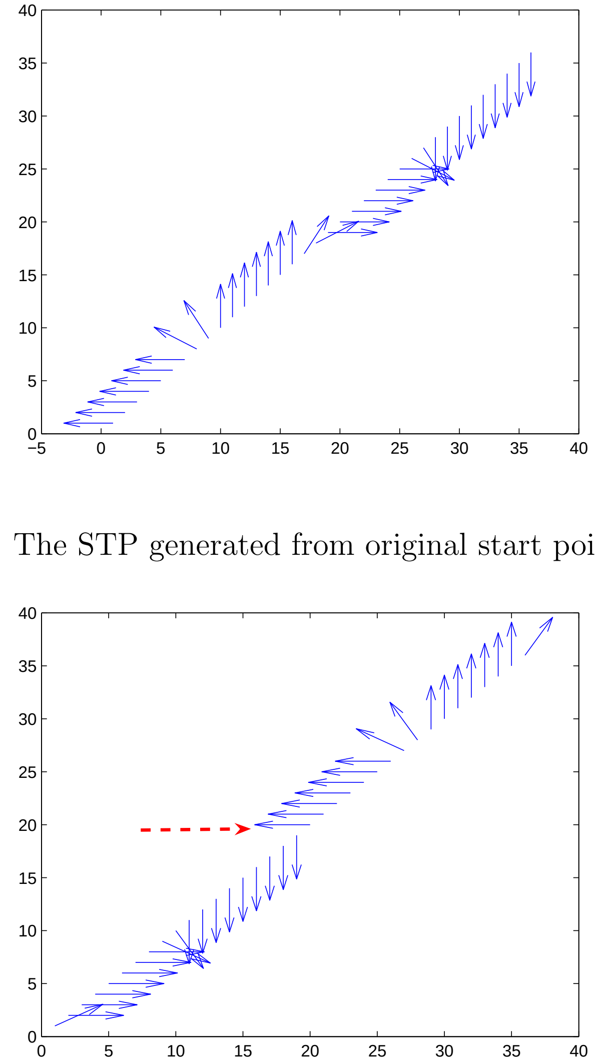


Fig. 16. **The 2D view of the STP for a square with smoothing.** The two arrows point to the two different locations where the STP sequences may have started.

of start locations for each STP. Having different start points is an important variation because the way the memory representation system was originally setup, all the STPs were generated by having the agent start at the same relative location on the shape and as a consequence the STP generated would not have much variation. Fig. 16 shows a 2D view of the STP for a square and possible positions where the STP sequence may start. Fig. 17 shows the corresponding 1D view.

The overall effect of adding different start points is that the average classification rate decreased with increasing number of start points (shown in Fig. 18). However, even with a high variation in the possible start points for starting STP the average classification rate was still higher than visual memory. These values were averages taken by performing ten trials with varying training and test data and a constant smoothing window size of three. All differences were significant under t-test ($p < 0.00015$, n = 10). In sum, skill memory was significantly easier to learn than visual memory, even when the task was made harder for STP. Note that the performance would suffer greatly if STP generation can start from an arbitrary point on the shape.



(a) The STP generated from original start point.

(b) The STP for a square shape with differing start point.

Fig. 17. **The linear ordering of the spatio-temporal pattern for a square with differing start points.** The plot in (a) shows how the STP will appear if the STP generation started from the original start point. The plot in (b) shows how the STP will appear if the STP generation started from the new start point (shown in Fig. 16). The dashed arrow in (b) points to the original starting location. The only difference with the original version is that the STP is shifted, but that is enough to affect the classification accuracy.

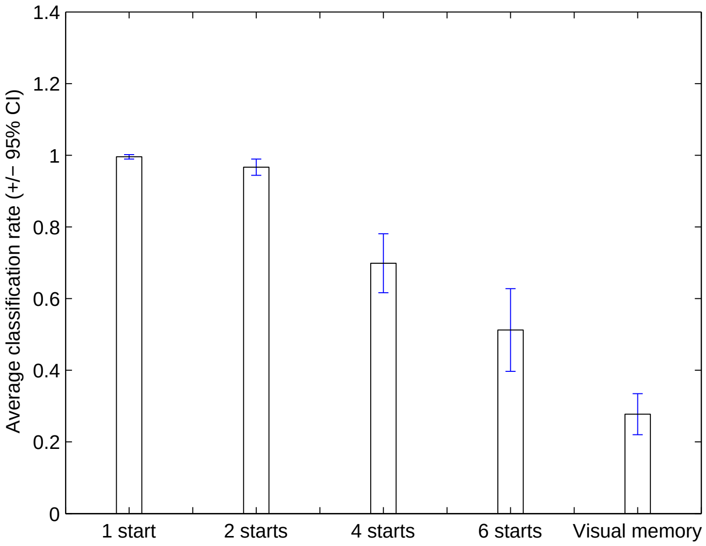


Fig. 18. **The effect of increasing start points on the classification rate for skill memory.** This bar chart shows the change in the average classification rate as the number of start points is increased (± 95 % CI). As the number of random start points is increased, the average classification rate steadily drops, but clearly skill memory performs better than visual memory.

However, in practical applications the STPs can be aligned using crosscorrelation analysis. Here, the purpose was mainly to show how skill-based memory performs under the most basic setup.

3. Trajectory noise

The last variation tried was the introduction of noise in the motion trajectory. This means that a random error at some point in time during the creation of the action sequence occurred causing the trajectory to deviate from its normal course. The action sequences were generated as before. However, at random an angle between 0 and 360 was added based on the magnitude of the noise factor. An example of noise in the trajectory is shown in Fig. 19 and Fig. 20. In these figures a noise factor of 0.1

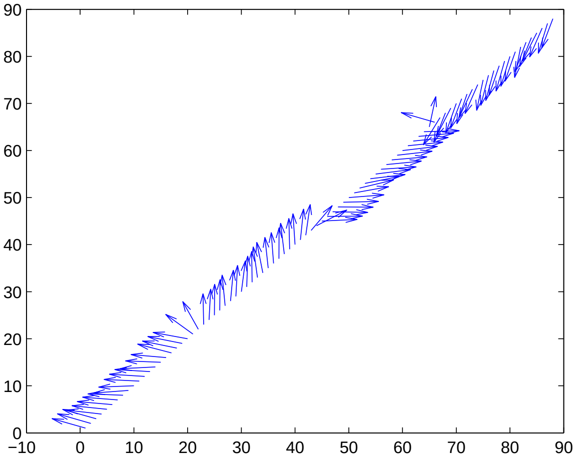


Fig. 19. **The STP with noise for a square shape**. The plot shows a 1 dimensional view of the spatio-temporal pattern for a square after the application of random noise (noise factor 0.1) and after smoothing.

was used.

The noise factor is the probability that affects the magnitude by which an action vector's angle in space may be affected. Hence, a larger noise factor will mean a larger deformation of the shape that a particular STP may trace. This angular change can range between 0 and 360. This range of angular change creates a problem for normalization of the angles which have to lie between 0 and 360 and then they have to be converted to lie between 0 and 1 by dividing all angles by 360 before being fed to the neural network. Since the change is positive in the range of 0 to 360. The angles larger than 360 were converted to lie between 0 and 360 by taking the modulus 360 of the angle. Fig. 21 shows how the classification rate decreases with the increase in noise. However, as we can see in Fig. 21 even at larger noise levels skill memory is still able to outperform visual memory. This demonstrates that skill-based memory is resilient to noise. All differences were significant under t-test ($p < 0.002$, n = 10).

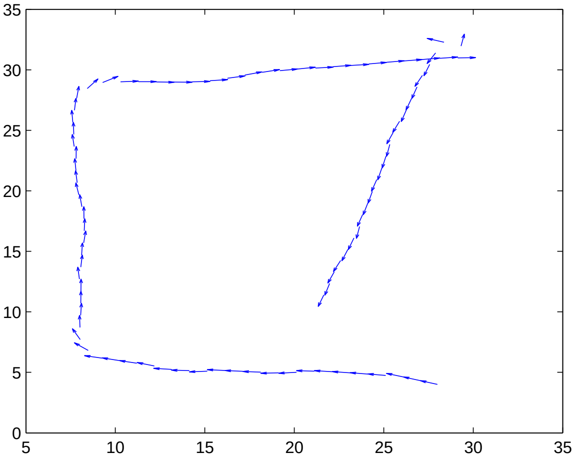


Fig. 20. **The STP with noise for a square shape**. The plot shows a 2 dimensional view of the spatio-temporal pattern for a square after the application of random noise (noise factor 0.1) and after smoothing. Notice how the shape traced out is still quite close to the original square.

C. Action as an intermediate representation

In order to test the hypothesis that action may serve as a good intermediate representation of sensory information, the following test was devised. The visual representation would be mapped to action sequence as well as the action sequence to action sequence mapping along with the visual representation to action. If the learning for visual to action is easier with respect to sensory to sensory mapping (e.g. visual to visual), then that would indicate that in fact sensory information can be easily represented in terms of action. This idea coupled with the primary view that action-based memory may perform better at object recognition tasks further supports the idea that skill memory be a more ideal form of memory. The reason for this being that if the sensory to action mapping was very difficult then the performance advantage that skill memory holds may become less pronounced and there by limiting the applicability of action-based memory.

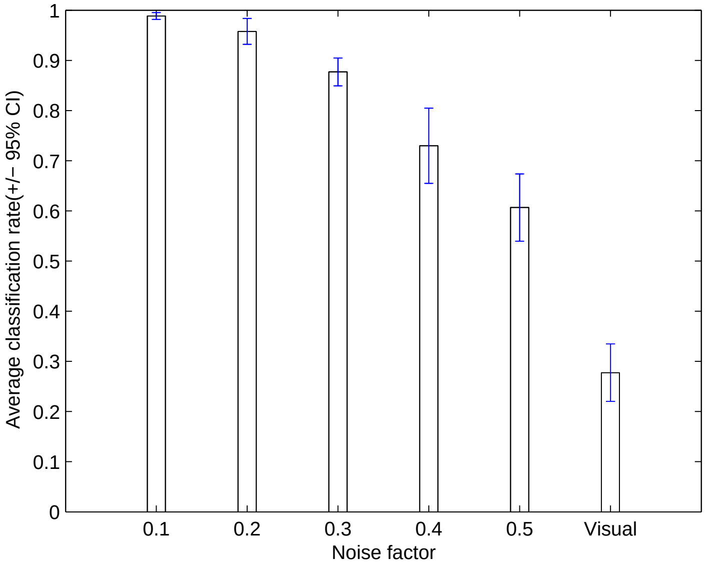


Fig. 21. **The effect of noise on classification rate.** This bar chart shows the effect of increasing noise on the average classification rate of skill memory (± 95 % CI). Visual memory is shown here as a baseline. The detrimental effect of noise can clearly be observed from this bar chart. However, notice that skill memory is quite resilient to noise and outperforms visual memory at noise factor of 0.5.

To verify this tests involving the different mappings were run. The different mappings were; action to action, visual to action, action to visual, and visual to visual. This test was carried out on a neural network with 900 inputs and 900 outputs, corresponding to the 900 dimensional input for each representation. Fig. 22 shows the results of the experiment. The figure shows that the learning curve for visual to action mapping is as low as action to action mapping (p = 0.37, n = 10). The figure also shows that the action to visual memory is slightly easier to learn than visual to visual mapping (however, t-test showed that $p = 0.82$, n = 10, indicating that the difference was not significant). All other differences were significant under t-test ($p < 0.026$, n = 10). This bolsters our idea that action may be a good intermediate representation for sensory data. This support makes action-based memory more appealing.

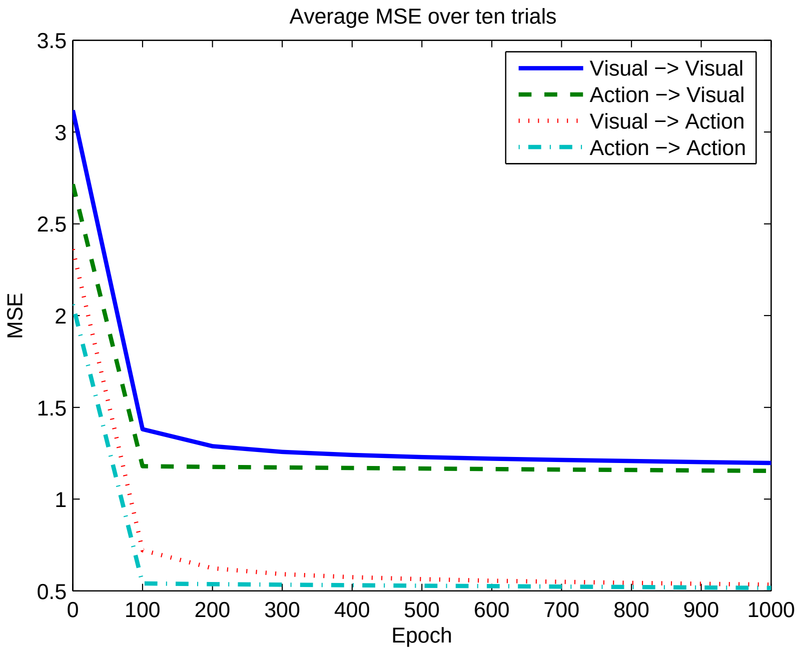


Fig. 22. **The learning curve for the different mappings.** This plot shows the learning curve for each mapping. From the learning curves we can infer how well each mapping performs with respect to other mappings. As expected the visual to action mapping performs well. The visual to action mapping is also relatively good.

CHAPTER V

DISCUSSION

Analysis of the results in the previous chapter clearly indicates that skill-based memory representation performs better than visual memory in recognition tasks. It has been additionally demonstrated that even under quite severe variations skill-based memory is able to yield results which indicates its merits. It has been further demonstrated that action may serve as a good intermediate representation for sensory information.

A. Properties of STP

The primary reason why skill memory yields such impressive results is because of its ability to capture the core discriminating property of the respective shapes. This is the case because aspects such as size and location of the figure do not cause variations in the resized STP. Hence the STP for different sized shapes was similar in the end. This makes it easy for the neural network to learn the skill-based representation. However, variations introduced to the STP to compensate for the apparent advantage, i.e., using noise and random start methods did not affect the results. Even large variations did not cause visual memory to out perform skill-based memory.

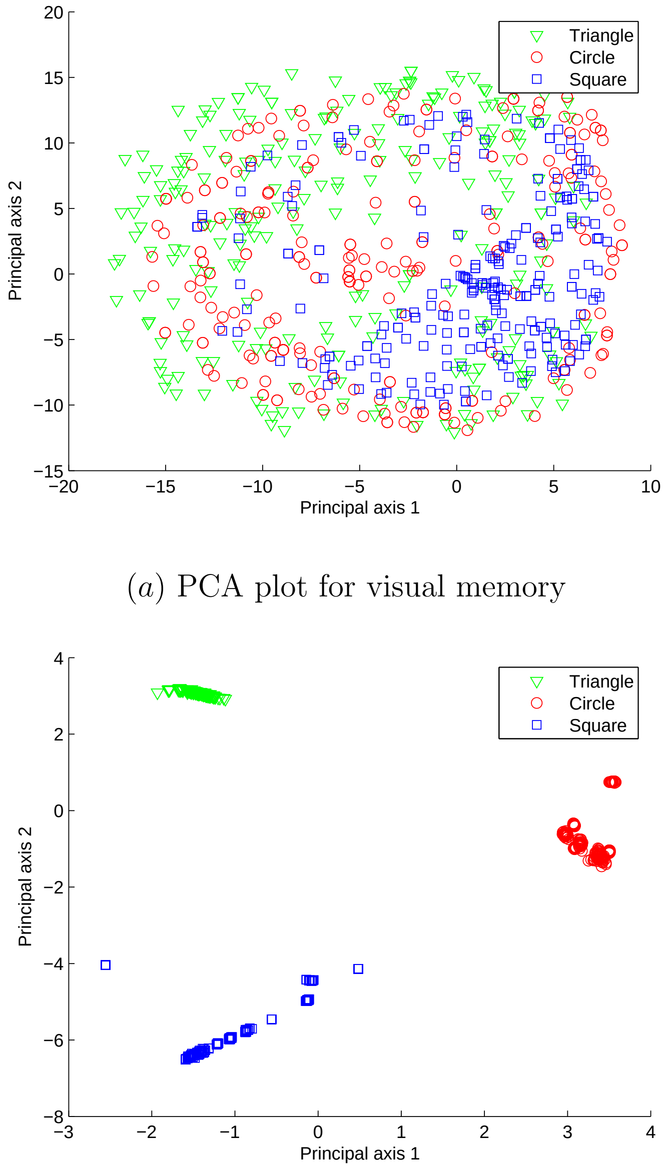
The properties of the STP and the 2D array representations can be represented as in Fig. 23. In this figure, the Principal Components Analysis (PCA) plots along two principal component axes are shown for the data points in the STP and the visual representations. Fig. 23(b) shows that the PCA plot of skill-based representation has three distinct clusters for the three classes of input shapes. On the other hand, the PCA plot for visual memory (Fig. 23(a)) has all the data points almost uniformly scattered, indicating that making proper class distinctions may be difficult. Such an analysis provides some insight on why skill memory performs better than visual memory in object recognition tasks.

Fig. 24 shows that with the introduction of noise the class boundaries become less pronounced. With low noise the three distinct clusters for the corresponding classes are maintained. As the noise factor is increased the clusters become less compact and it becomes slightly harder to determine the class boundaries. However, even with high noise the class boundaries can still more or less be determined. These plots help us understand why the neural network was marginally less able to properly recognize skill-based representations when the noise was high.

With the introduction of varying start points many more clusters appear in the PCA plot, as illustrated in Fig. 25. However, it is interesting to note that the local clusters are more compact as opposed to the broader clusters that emerge with the addition of noise in Fig. 24. That is, data points from the same class are scattered around but they locally form tight non-overlapping clusters.

B. Potential limitations

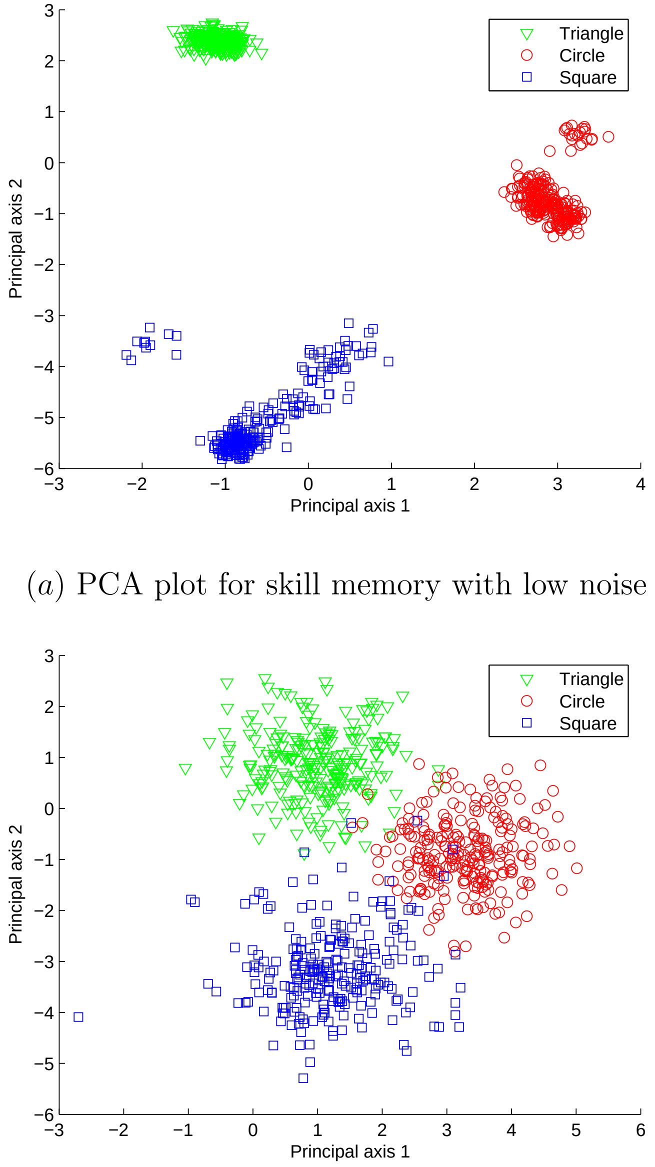
One of the main assumptions of this research is that action can be represented along a time series that are scaled to be of the same length. However, one may question the validity of creating an action sequence and scaling such an action sequence. This also leads one to question, at what intervals are the actions stored and should this interval be long or short. All of these questions can be answered by the recent experiments performed by Conditt et al. [12]. The result of their experiment suggests that when humans are asked to perform a series of actions, the actions tend to be represented as time invariant. This means that humans do not store the actions parameterized by time. More precisely, humans do not have a timing mechanism that stores the



(a) PCA plot for visual memory

(b) PCA plot for skill memory

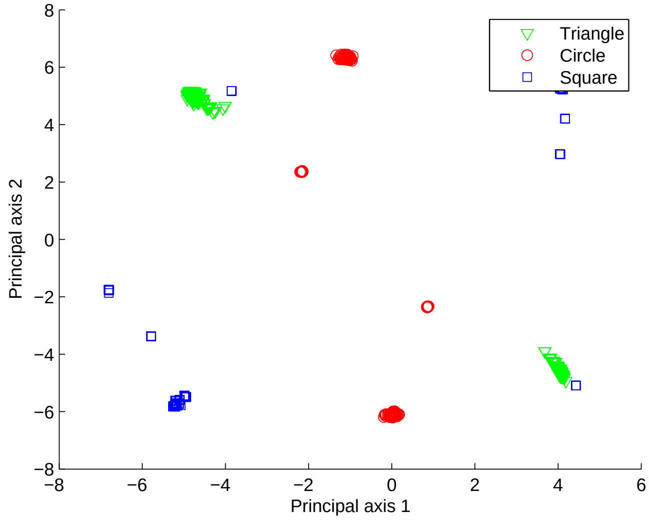
Fig. 23. **The plot of PCA projection for visual and skill memory.** (a) PCA for 2D array representation (visual memory) of the input data along the first two principal axes is shown. (b) PCA for STP representation (skill memory) of the input data along the first two principal axes is shown.



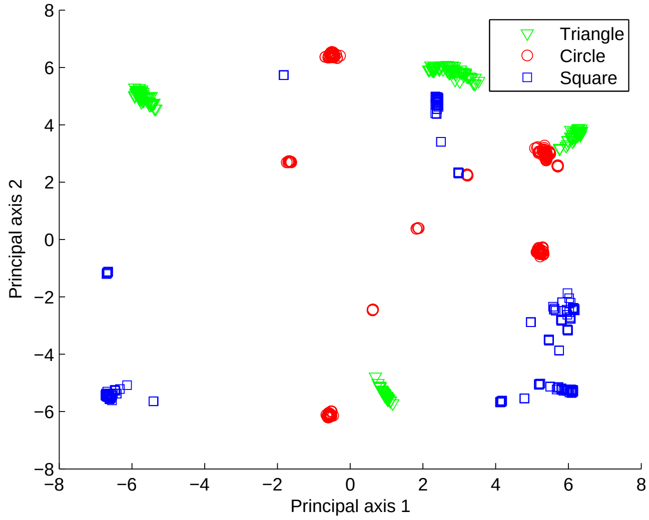
(*a*) PCA plot for skill memory with low noise

(*b*) PCA plot for skill memory with high noise

Fig. 24. **The projection on the two principal axes for skill memory with varying noise factor.** (*a*) PCA for skill memory with a lower noise factor of 0.1 along the first two principal axes is shown. (*b*) PCA for skill memory with a high noise factor of 0.5 along the first two principal axes is shown.



(*a*) PCA plot for skill memory with few start points



(*b*) PCA plot for skill memory with many start points

Fig. 25. **The projection on the two principal axes for skill memory with varying number of random start points.** (*a*) PCA for skill memory with two random start points for STP generation along the first two principal axes is shown. (*b*) PCA for skill memory with four random start points for STP generation along the first two principal axes is shown.

exact time period between actions. This form of representation allows humans to counteract disturbances in the environment. A disturbance can result in the delay in the completion of an action for a given sequence. However, most humans are able to go along and complete the rest of action. Such experimental evidence allows us to be more confident about representing actions in a spatio-temporal pattern in the way I did in this thesis.

The STPs produced in these experiments seem to be the same for each shape, thus making the job of learning trivial and therefore it appears that skill memory had an unfair advantage. Rather than this be a criticism against the validity of the research, it points out the fact that STP is not affected by size and translation of object. The similarity in STPs further point out the core thesis of the research, that action sequence as represented in skill memory may be an inherently superior representation scheme than the raw sensory information as in visual memory, because of its ability to capture properties of the object when time can be scaled with ease.

C. Relation to memory in humans

The two memories, visual and skill, are analogous in many ways to the types of memory employed by natural agents. Natural agents have episodic and procedural memory [13]. Episodic memory is fact-based where certain information about events is stored. It is currently believed that these events are temporarily stored in the hippocampus [14] [3]. This may have similarities to visual memory described above. On the other hand, skill memory can be thought of as being similar to procedural memory. Procedural memory deals with the ability to recall sequential steps required for a particular task that an agent may perform [15]. It may be interesting to investigate if the main results derived in this thesis applies to the understanding of human memory and recognition.

Note, however, that I am not trying to claim that the results presented here explain how the human memory works. Rather, what is presented here only suggests that skill-based memory may have theoretical virtue regarding perceptual understanding, as compared to sensory memory.

D. Future work

The future expansion of this research topic involves the actual implementation of the skill-based memory system in an autonomous agent such as SIDA. Also other variations to the STP format can be studied such as methods that retain only the changes in the sequence of actions, i.e. only when there is a certain change in action, rather than retaining the total sequence. The resultant STP for shapes like square will have only four points, since in the traversal of a square the action vectors will need to be only changed four times.

In order to see if visual memory can perform better at object recognition, it might be a good idea to provide a mechanism that lowers the dimensionality of the sensory data. This can be done by splitting the visual field into smaller patches and capturing the activation in the smaller visual area. Since the number of dimensions is reduced it might be slightly easier for the neural network to learn the visual memory.

E. Contributions

The primary contribution of this research is the demonstration that skill-based memory has beneficial properties that can aid in perceptual understanding. These properties are in line with other research that suggested that action is a fundamental component for learning simple properties. However, in this thesis I was able to demonstrate that action plays an important role in learning complex objects when the system was allowed to have memory. This research clearly demonstrates how action can be incorporated into a powerful autonomous learning system.

CHAPTER VI

CONCLUSION

This thesis, the study of memory systems, arose from the desire to develop a memory system that would allow autonomous agents to learn about complex object properties. The most basic memory system that an agent can have is the direct (raw) storage of the sensory data (such as visual memory). Another system is skill-based memory, which primarily involves the retaining of action sequences performed during a task. Skill memory was anticipated to be a better representation because of the crucial role action played in simple perceptual understanding [1]. To test this hypothesis, I compared the two memory representations in object recognition tasks. The two primary performance measures, average classification rate and MSE, revealed the superior properties of skill memory in recognizing objects. Additionally, a related experiment demonstrated convincingly that action can serve as a good intermediate representation for sensory data. This result provides support for the idea that suggests that various sensory modalities may be represented in terms of action.

Based on the above results, it can be concluded that the importance of action in simple perceptual understanding of objects can successfully be extended to that of more complex objects when some form of memory capability is included. In the future, the understanding we gained here is expected to help us build memory systems that are based on the dynamics of action that enable intrinsic perceptual understanding.

REFERENCES

[1] Y. Choe and S. K. Bhamidipati, “Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action,” in *Biologically Inspired Approaches to Advanced Information Technology*, ser. Lecture Notes in Computer Science 3141, A. J. Ijspeert, M. Murata, and N. Wakamiya, Eds. Berlin: Springer, 2004, pp. 176–188. [Online]. Available: http://faculty.cs.tamu.edu/choe/ftp/publications/choe.bioadit04.pdf

[2] L. R. Squire, “Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans.” *Psychological Review*, vol. 99, pp. 195–231, 1992.

[3] R. L. Buckner, “Neural origins of ’i remember’,” *Nature Neuroscience*, vol. 3, pp. 1068–1069, 2000.

[4] D. Philipona, J. K. O’Regan, and J.-P. Nadal, “Is there something out there? Inferring space from sensorimotor dependencies,” *Neural Computation*, vol. 15, pp. 2029–2050, 2003. [Online]. Available: http://nivea.psycho.univ-paris5.fr/Philipona/space.pdf

[5] J. K. O’Regan and A. Noë, “A sensorimotor account of vision and visual consciousness,” *Behavioral and Brain Sciences*, vol. 24(5), pp. 883–917, 2001. [Online]. Available: http://www.bbsonline.org/Preprints/ORegan/

[6] J. C. Buisson, “A rhythm recognition computer program to advocate interactivist perception,” *Cognitive Science*, vol. 28, pp. 75–87, 2004.

[7] I. W. J. Y. Aloimonos and A. Bandopadhay, “Active vision,” *International Journal on Computer Vision*, vol. 1, pp. 333–356, 1988.

[8] K. S. Lashley, “The problem of serial order in behavior,” in *Cerebral Mechanisms in Behavior*, L. A. Jeffress, Ed., Wiley, New York, 1951, pp. 112–146.

[9] H. Abelson, *LOGO for the Apple II.* Peterborough, N.H.: McGraw-Hill.

[10] S. Haykin, *Neural Networks: A Comprehensive Foundation.* Englewood Cliffs, NJ: Macmillan, 1994.

[11] J. A. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley, 1991.

[12] M. A. Conditt and F. A. Mussa-Ivaldi, “Central representation of time during motor learning,” *Journal of Neurobiology*, vol. 20, pp. 11 625–11 630, 1999.

[13] Y. Silberman, R. Miikkulainen, and S. Bentin, “Semantic effect on episodic associations,” *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland, vol. 23, pp. 934–939, 1996.

[14] E. Tulving and H. J. Markowitsch, “Episodic and declarative memory: role of the hippocampus,” *Hippocampus*, vol. 8, pp. 198–204, 1996.

[15] S. P. Wise, “The Role of the basal ganglia in procedural memory,” *Seminars in Neuroscience*, vol. 8, pp. 39–46, 1996.

VITA

Navendu Misra was born in India in 1981. At the age of nine he moved to Trinidad and Tobago. He completed his GCE O'Levels in 1997 and GCE A'Levels in 1999 with distinctions while attending Presentation College San Fernando. Upon completion of his secondary education, he went on to attain his Bachelor of Science degree in computer science from The University of Texas at Austin in May 2003 with university honors. He is expected to receive his Master of Science degree in computer science at Texas A&M University, College Station, in August 2005.

Permanent Address: 123 Wittet Drive, Central Park, Balmain, Couva, Trinidad, West Indies.