〔非公開〕

ＴＲ－Ｃ－０１５８

Ｒｅｃｏｇｎｉｔｉｏｎ ｏｆ Ｃｏｎｔｉｎｕｏｕｓ Ｇｅｓｔｕｒｅｓ Ｕｓｉｎｇ Ｎｏｎｌｉｎｅａｒ Ｄｙｎａｍｉｃｓ

エドワード アルトマン
Edward ALTMAN

１９９６ ３．１５

ＡＴＲ通信システム研究所

Recognition of Continuous Gestures Using Nonlinear Dynamics

Edward J. Altman

ABSTRACT

Early gesture recognition systems used discrete gestures for pointing and grasping tasks. Later systems emphasized the use of continuous gestures and sign language gestures which contain more abstract information. The thesis of this paper is that the temporal qualities of continuous gestures can be modeled more naturally using dynamical systems. The primary problem is the construction of model dynamics for gestures. A secondary problem which arises from continuous gestures is the temporal alignment, segmentation. and classification of the hand motions. The goal of this research is to exploit the multifunctionality of nonlinear systems to develop a concise mechanism for gesture recognition.

This paper describes a nonlinear technique for gesture recognition based on selective synchronization in a population of dynamical systems. A genetic algorithm is developed to learn the synchronizing dynamics. The emergent property of spatial pattern formation in a population of dynamical systems is used to perform classification. The advantages and disadvantages of nonlinear dynamics for gesture recognition are discussed and future directions for this research are suggested at the end of the paper.

1 Introduction

Many natural phenomena exhibit distinct temporal rhythms and rapid transitions between rhythms due to external interactions[1]. The ability to model these behaviors is often an important component of a recognition system. An everyday occurence for which rhythmic motion is important is the recognition of simple hand gestures motivated by American Sign Language (ASL)[2]. Hand gestures in ASL have a high degree of spatio-temporal structure which can be modeled as motion on a manifold surface. Transitions between gestures then correspond to transitions between manifolds. Recognition requires; 1) the temporal alignment of the model to the input signal, 2) segmentation of meaningful gestures from the input stream, and 3) classification of the trajectory. In this paper, synchronization within an array of dynamical systems provides a mechanism for the concurrent alignment, segmentation, and classification of a simple test set of gestures.

A hand gesture is a motion in space with characteristics of speed, spatial position, tension, hand shape, co-articulation, and other possible dimensions[2]. Gestures may occur in an unlimited number of combinations and shapes, therefore a generic form of interaction is required based upon selective synchronization, topological structure, and bifurcations of nonlinear dynamical systems. Since the gesture is dynamic, it has a natural encoding by dynamical systems. Synchronization among dynamical systems at the levels of an individual system, a cluster of neighboring systems, and globally throughout the array correspond to temporal alignment, segmentation, and classification, respectively.

A gesture may begin with an arbitrary phase due to context dependent transitions between gestures. Therefore, at the level of individual dynamical systems, synchronization on an attracting manifold provides a selective mechanism for temporal alignment of the model dynamics to the input signal. A learning method based on a genetic algorithm is used to acquire the synchronizing dynamics for the trajectory of a sample gesture.

Since gestures in ASL have spatio-temporal structure, this structure is represented topologically as a manifold surface. Dynamical systems with similar attracting manifolds will tend to become synchronized to each other when the input trajectory lies near the manifold surface. The coherence of oscillations among neighboring elements in an array provides a local amplification of the responses by the individual elements responding to a learned trajectory. For a trajectory which does not lie near the attracting manifold, the individual differences among dynamical systems result in a rapid divergence of trajectories. Thus, local averaging of responses increases robustness and reduces dependence on initial state.

The gestures in ASL are high dimensional signals which vary according to position, speed, tension, hand shape and co-articulation. In order to simplify the gesture recognition problem, nonlinear models have been constructed only for the position trajectories of the hand motion. Gesture trajectories interact with the dynamical systems by means of diffusive coupling and result in selective synchronization. Spatially localized coherency develops among similar models in a 2D array of model dynamics responding to the same input. This coherency at a local level generates

global patterns of synchronization which provide relative information about multiple features detected in the input signal.

The goal of this research is to exploit the multifunctionality of nonlinear systems to develop a concise mechanism for gesture recognition. Primary emphasis is placed on determining which properties of nonlinear systems can be used to perform the temporal alignment, segmentation, and classification of the gesture data. The efficient implementation of a gesture recognition system based on nonlinear mechanisms is a task for future research.

The remainder of this report is organized as follows. Section 2 describes the synchronization mechanism for gesture recognition. In Section 3, the nonlinear dynamics paradigm is discussed in terms of the efficient design of analog devices based upon the multifunctionality of nonlinear systems. Based upon the experiences of designing the gesture recognition array, the key advantages and disadvantages of the monolithic design of complex dynamics is discussed. Directions for future research are discussed in Section 4.

2 Mechanism for Gesture Recognition

In this paper, we describe computational modules based upon the process of synchronization in an array of dynamical systems at three levels. At the level of an individual system in Fig. 1, synchronization provides the temporal alignment of the system to the input trajectory. At the local level segmentation of the input is performed by the synchronization among neighboring systems resulting in the coherency of oscillations among systems with similar dynamics. Globally, the selective synchronization and local coherence give rise to pattern formation in Fig. 1. The slowly changing spatial pattern is interpreted as an energy field which determines the low dimensional output dynamics of the 2D array. Alternatively, a neural network classifier can be used to perform gesture recognition in the transformed domain of spatial patterns. The ensemble of dynamical systems which concurrently performs these three tasks is called the **gesture recognition module**.

2.1 Synchronization

The trajectories of hand gestures with sufficient spatial structure can be modeled as motion on a manifold surface. The construction of models from natural systems is, in general, a complex problem. Therefore we use a genetic algorithm [3] as a learning method for searching a high dimensional parameter space for the coefficients of the synchronizing dynamics. In this section, a general system of the form

$$\begin{aligned} \dot{x} &= \sigma(f(x,y,z)) + \alpha_1(F_x - x) \\ \dot{y} &= \sigma(g(x,y,z)) + \alpha_2(F_y - y) \\ \dot{z} &= \sigma(h(x,y,z)) + \alpha_3(F_z - z) \end{aligned} \tag{1}$$

where f, g and h are nonlinear functions containing terms of degree ≤ 2, $\sigma(\zeta) = (1 + e^{-\zeta})^{-1}$ is the standard sigmoidal function, and α_i is the diffusive coupling

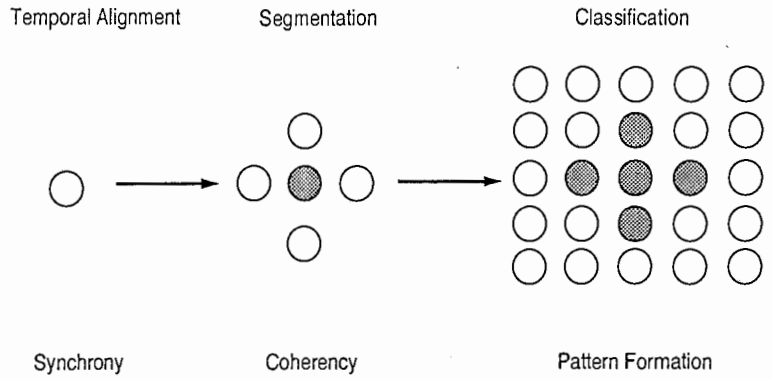

Figure 1: Levels of interactions in an array of dynamical systems.

coefficient. Although the variables x, y, and z may be associated with any abstract properties, we shall use a direct mapping of spatial coordinates. Specifically, the external forcing term F_x corresponds to the x-component of the input trajectory. The term $\alpha(F_x - x)$ describes the coupling between the input and the dynamical system. Here we use a simple method of interaction based on diffusive coupling.

The parameters of Eq. (1) define the space of a dynamical system. A genetic algorithm, *Genocop*[3], was used to search the parameter space of the functions $f(x, y, z)$, $g(x, y, z)$, and $h(x, y, z)$ for a dynamical system which synchronizes to a sample trajectory. During learning, the gesture trajectory, $\mathbf{F}$, is coupled to the dynamics of Eq. (1), and the fitness function evaluates the synchronization among all components x, y, and z of the dynamical system. The fitness function for synchronization is defined as

$$fitness = \sum_{n=1}^{N} \mid \mathbf{x}_n - \mathbf{F}_n \mid \tag{2}$$

which is the Euclidean distance between the 3D input, $\mathbf{F} = (F_x, F_y, F_z)$, and the state of the dynamical system $\mathbf{x} = (x, y, z)$ summed over the length of the trajectory. Notice that learning utilizes all dimensions of the input data, whereas recognition may be performed on a lower dimensional projection of the input. In this case, only the x-component is required for synchronization since the input is assumed to be on a manifold surface.

The autonomous form of the learned dynamics is obtained by setting $\alpha = 0$ in Eq. (1). The autonomous dynamics exhibits a fixed point attractor near the center of the manifold for the learned gesture. One constraint on the autonomous system is that all initial conditions associated with valid states of the input trajectory should lie within the attracting domain of the dynamical system. The input trajectory is shown in Fig. 2 as a dashed curve. Each sample point is used as the initial condition for Eq. (1) and converges onto the fixed point attractor. Therefore the autonomous system has an extremely simple behavior. Moreover, there is no explicit representation of the learned trajectory.

The representation of the learned gesture is determined dynamically through the

interactions between Eq. (1) and the input trajectory. When the input $\mathbf{F}$ is applied to Eq. (1) for $\alpha \neq 0$, the system bifurcates into a series of fixed points which divide the manifold into two sections shown in Fig. 3. In this case, the locations of the fixed points depend upon the external driving from $\mathbf{F}$. The effect of the external forcing is to convert the static attractor into an attractor parameterized by time. For each successive point along the learned trajectory, the attractor is moved along a curve to a position such that the short term integration of Eq. (1) follows a short segment of the input trajectory. The cummulative effect of this interaction is the synchronization of the dynamics to the input. This shows that the functionality of the learned dynamics is determined by interactions with the input.

The synchronization mechanism for the case of 1D driving from F_x for the periodic motions given above can clearly be described interms of the geometry of the attractor. Similar mechanisms are observed for the learning of 3D inputs for gestures from ASL. Figure 4 shows the synchronization of the learned dynamics to the gesture combination meaning "have book". The highlighted points in Figure 4 are shown in more detail in Figure 5. In this case we see that the short term temporal integration of the driven dynamics results follows the input trajectory. Successive inputs then change the path of the integration to result in synchronization.

2.2 Coherence

The architecture for the recognition system consists of multiple levels, each of which extracts particular information and relations from the gesture stream. The nonlinear phenomenon of synchronization provides the temporal alignment between the model dynamics and the input data near the learned manifold surface. The coherence of similar dynamical systems responding to the same input on the manifold surface provides an indirect method for segmentation.

During the gesture learning phase, multiple samples of the same gesture are used to construct multiple models using synchronizing dynamics. Since the input data near the manifold surface of the gesture is similar in each sample, the dynamics of Eq. (1) near the manifold surface of the gesture data should be the same. The sensitive dependence of the dynamics on its parameters insures that no two dynamical systems will have the same response away from the manifold surface. The use of population coding with convergence near the manifold surface and divergence away from the manifold provides a mechanism for coherence based segmentation.

2.3 Pattern Formation

At the third level of concurrent processing the local coherency in the gesture array gives rise to global patterns of synchronization. The gesture array is constructed from subarrays of dynamical systems which are trained to synchronize to specific learned trajectories. The purpose of the recognition array is to convert the temporal signal from the hand gesture into a static 2D image which is simpler to recognize. Spatial pattern formation simplifies recogniton by performing the task of dimensionality reduction.

The recognition array is composed of subarrays of synchronizing dynamical systems which can be trained to variants of the same gesture in order to improve the robustness of the system to external noise. The subarrays also serve to compensate for the dependence of the dynamical systems upon the initial state. Since vector field of nonautonomous nonlinear systems is complex, the stability of the system to a range of inputs and initial states cannot be tested exhaustively. Therefore the scheme of population coding minimizes the effects of the sensitivity to initial conditions.

3 The Nonlinear Dynamics Paradigm

Many problems in nature can not be adequately modeled by linear processes. Any recognition task is inherently nonlinear since a differential response to simlar inputs is required to assign a classification to the input. It is the "great promise" of nonlinear systems that some particularly hard problems may have relatively simple solutions, provided the properties on nonlinear systems can be successfully harnessed. The focus of the present research on gesture recognition has been; 1) learning of synchronizing dynamics as models for gestures, and 2) constructing a monolithic system of nonlinear dynamics which achieves a high degree of multifunctionality. This section discusses the insights gained from this research.

3.1 Advantages of the Nonlinear Approach

The multifunctionality of nonlinear systems is a consequence of bifurcation phenomena and emergent behavior. Bifurcation phenomena give raise to distinct responses in the model dynamics as a consequence of parameter changes. Emergent behavior arises from the occurrence of similar behaviors over spatial or temporal scales. Multifunctionality offers the following advantages for nonlinear systems:

1. Reduction of multiple processing stages to a single processing step.

2. Possibility of modeling emergent behavior in complex systems.

3. Utilization of the special characteristics of nonlinear phenomena.

The analog implementation of complex, nonlinear systems relies upon the efficient design of the dynamics. The abovementioned advantages will next be examined for their potential contribution to a hardware implementation.

Recent advances in analog VLSI have for the first time made feasible the implementation of large arrays of dynamical systems[4]. In spatially distributed analog systems, the communication among multiple processing stages creates a severe hardware interconnection problem. Therefore the multifunctionality of nonlinear systems may help reduce the effects of this problem by performing more complex processing tasks within a single monolithic system. In the case of the gesture recognition array, the concurrent temporal alignment, segmentation, and classification within a

single array of dynamical systems illustrates the complex processing capabilities of nonlinear systems.

Recently, there has been much interest in the exploitation of emergent behavior in complex systems to achieve a new computational paradigm. Research on spatially distributed systems has shown that simple nonlinear systems arranged into a uniform latice can produce a rich varity of behaviors through local interactions. The selective synchronization of nonlinear systems to continuous inputs has been shown to result in the emergent property of pattern formation[5]. In this case the dynamical systems are not uniform throughout the array. The learning proceedure insures that neighboring systems in the array have similar synchronization behavior in the neighborhood of the learned manifold surface of the target gesture. However, the differences between neighboring systems when situated away from the learned manifold surface insures the divergence of behaviors for non-target input trajectories. The emergent behavior of patterns of synchronization therefore arise from the sensitive dependence of nonlinear systems away from the attracting manifold surface and their invariance near the manifold.

By addressing the problem of multifunctionality, it is possible to demonstrate the utilization of the special characteristics of nonlinear phenomena. The role of bifurcations between two attracting states of simple nonlinear systems has been shown to determine the propagation of information in an array of uniform dynamical systems. In the gesture recognition array, the autonomous systems have a single attracting state, but nonautonomous systems responding to gestural inputs result in a continuum of attracting states such that synchronization occurs between the system and the input.

3.2 Difficulties of the Nonlinear Approach

In the course of constructing the recognition array several difficulties with the monolithic approach to gesture recognition became apparent.

1. Modeling the problem and measuring progress is uncertain due to a lack of mathematical theory.

2. Modifications to the implementation are difficult since multiple functions are interdependent.

3. Excessive demands are placed on the learning algorithm. A combination of dynamics design and refinement by learning is necessary.

4. Dimensionality of spatially extended systems is very high.

5. Dependence on diffusive coupling for synchronization does not allow for the learning of precise models.

Most of these difficulties arose from the ambitious attempt to perform the three separate tasks of temporal alignment, segmentation, and classification within a single monolithic system. A layered architecture with distinct modules and specific interactions between modules would eliminate many of the above problems.

Table 1: Levels of Processing for Gestures

classification	pattern formation induced from features
style	customization of dynamics
dynamic features	tension, rhythm, excentricity, etc.
motion filters	motion primitives
sensing	optical flow, region detection

4 Future Directions

Nonlinear dynamical systems offer considerable promise for performing complex functions with simple structures. However, practical applications of this emerging technology are difficult to achieve since formal methods of analysis and design are still in an early stage of development. The application of nonlinear dynamics for continuous gesture recognition has provided a framework for investigating the utilization of various properties of nonlinear systems, such as bifurcations, attracting manifolds, chaos, and synchronization. The particular emphasis of the current research was to determine the extent to which multifunctionality can be achieved in nonlinear systems by taking advantage of fortuitous commonalities in the behaviors of the nonlinear systems.

Future work on dynamics based gesture recognition should be redirected toward a behavior based decomposition of gestures according to the subsumption architecture[6]. The key layers of the decomposition are described in Table 1.

Each of the levels described in Table 1 may be implemented as nonlinear systems. Based on the experiences gained from the monolithic system, nonlinear dynamics can be used in the following ways:

1. synchronization to coordinate timing,
2. stable manifolds to represent spatio-temporal relationships,
3. bifurcations between alternative interpretations,
4. rapid transitions through state space due to chaos,
5. complex behaviors from simple systems.

The task of classification relies primarily upon; 1) manifold surfaces to represent possible configurations, 2) bifurcations between alternative interpretations, and 3) rapid transitions between stable states. The specification of gesture style relies upon the existence of complex behaviors from simple systems with a few parameters having specific meaning for the gesture, such as tension, excentricity, or velocity. The composition of dynamic features may require the use of synchronization to specify the relative timing of lower level features.

Whereas the previous research focused on integrating the tasks of temporal alignment, segmentation, and classification into a single recognition array, the direction of future research is toward building a layered architecture with interaction between

layers. This would lead to a modular structure which would simplify the application of nonlinear principles to the system design.

References

[1] T. Pavlidis, *Biological Oscillators: Their Mathematical Analysis.* New York: Academic Press, 1973.

[2] H. Poizner, E. S. Klima, and U. Bellugi, *What the Hands Reveal about the Brain.* Cambridge, Massachusetts.: The MIT Press, 1987.

[3] Z. Michalewicz, ed., *Genetic Algorithms + Data Structures = Evolution Programs.* New York: Springer-Verlag, 1992.

[4] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Trans. Circuits Syst.*, vol. 40, no. 3, pp. 147–156, 1993.

[5] E. J. Altman, "Hand trajectory recognition using dynamical systems," in *ACCV93 Asian Conference on Computer Vision*, (Osaka, Japan), pp. 321–324, Nov. 1993.

[6] R. Books, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. RA-2, pp. 14–23, 1986.

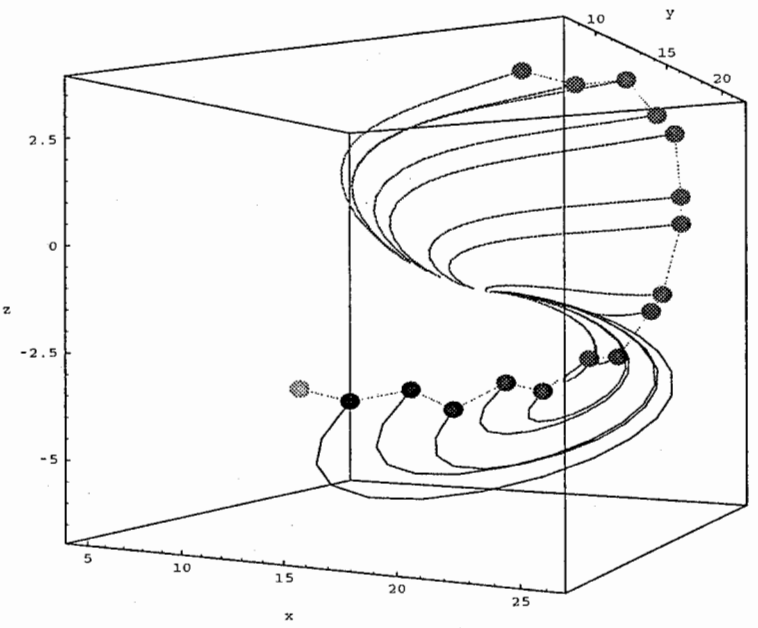

Figure 2: Convergence of all states from the input gesture to a fixed point attractor of the autonomous dynamics.

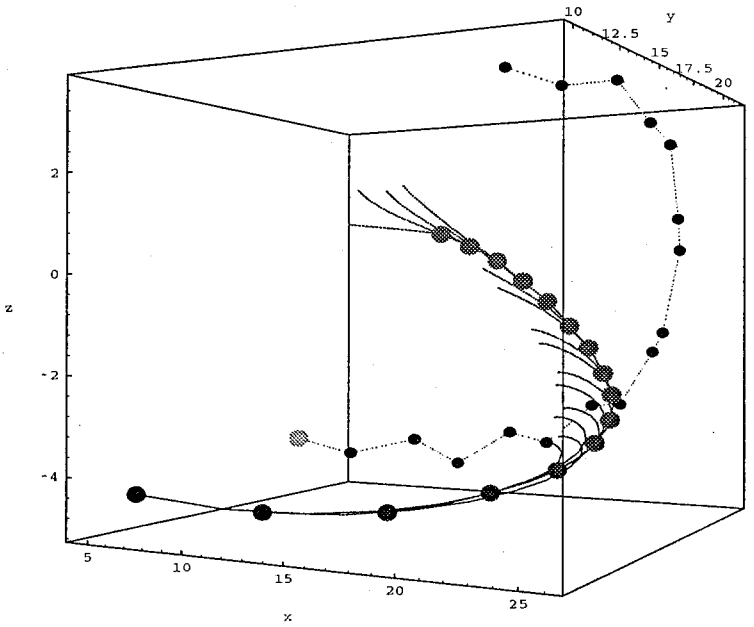

Figure 3: Mechanism of synchronization due to modification of the fixed point attractor by the input gesture.

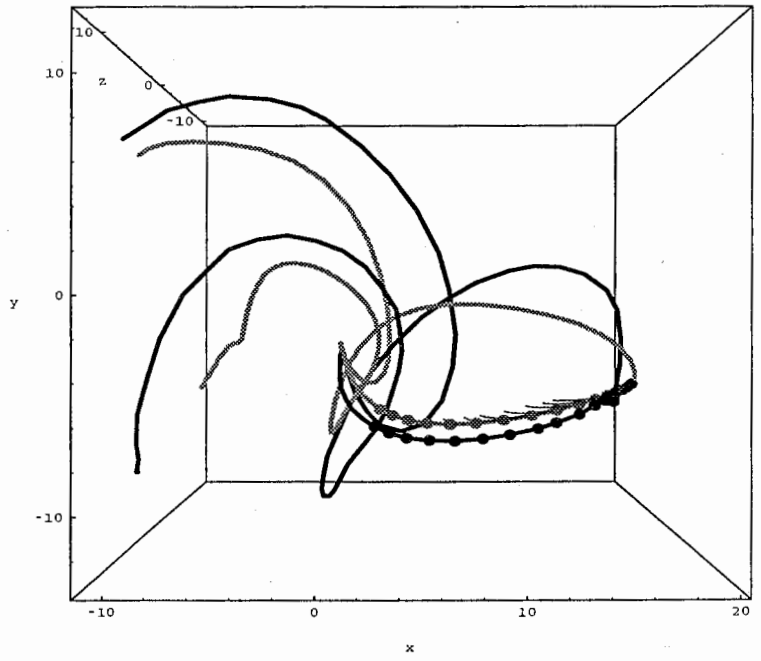

Figure 4: Synchronization of model dynamics (grey) to the gesture trajectory for the ASL phrase "have book". Forward integration is performed on the highlighted points to show the short term prediction of the model dynamics with external forcing.

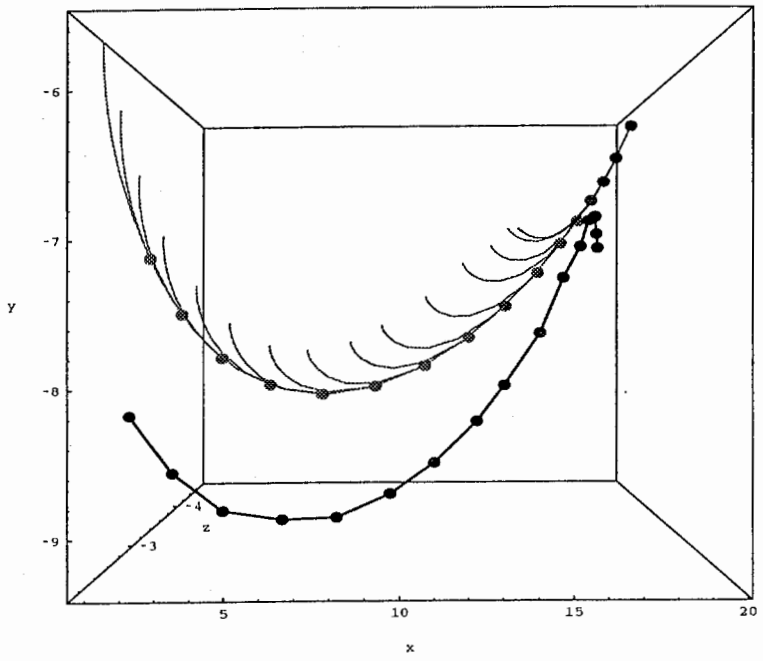

Figure 5: Expanded view of the forward integration of model dynamics. The stable attracting point of the model dynamics for the autonomous system is modified by the external forcing, thus resulting in synchronization to the learned input trajectory.