Neural Network Models in Evolutionary Ecology : How to Handle Optimal Reaction to Multiple Inputs

加茂, 将史
九州大学理学研究科生物学専攻

https://doi.org/10.11501/3180862

出版情報：九州大学, 2000, 博士（理学）, 課程博士
バージョン：
権利関係：

Neural Network Models in Evolutionary Ecology: How to Handle Optimal Reaction to Multiple Inputs

加　茂　将　史

Neural Network Models in Evolutionary Ecology: How to Handle Optimal Reaction to Multiple Inputs

by

Masashi Kamo

Department of Biology,

Faculty of Science,

Kyushu University

Fukuoka 812-8581, Japan

Dissertation

submitted in partial fulfillment of the requirements

for the degree of Doctor of Science

in Biology

Division of Science, Graduate School of Kyushu University

CONTENTS

PREFACE

The evolution of the form of signals used in animal communication has received much interest in recent years. The exaggerated traits in males used as sexually attractive ornaments have been studied especially intensively in the context of sexual selection. The traditional ideas for the evolution of these traits are roughly classified as two categories: Fisher's runaway theory and handicap principle theory. For example, suppose that females in a population have a small tendency to choose the males with a longer tail than those with shorter ones. A female with a stronger preference can obtain an indirect benefit from the high mating success of her sexy sons who inherit their father's long tail. Hence her genes causing a stronger choice than average females in the population will spread, together with the exaggerated tails of males, resulting in the evolution of spectacular tail of males (Fisher 1930; Pomiankowski *et al.* 1991). The second is handicap principle by Amot Zahavi (1977), who pointed out that females often use male's ornament as an indicator of "quality" of males. Provided that only a truly strong male with a high genetic quality can afford to produce and maintain the costly ornaments. This idea has been formalized in terms of quantitative genetics (Iwasa *et al.* 1991) and a signaling game (Grafen 1990).

A third idea has been proposed to explain the evolution of exaggerated traits. It suggests that the males' traits might have evolved through the preexisting bias in females' sensory mechanisms which have evolved in a context unrelated to the male's trait. For example, since females' sensory organ has a propensity to respond more to a stronger stimuli, such as a stronger voice, a louder call or a longer tail, males with these stimuli would be more likely to be chosen by females as a partner than the males emitting weaker stimuli. Females may simply react to a stronger stimulus, rather than they specifically "choose" males that can produce stronger stimuli as it indicates the quality of males. This idea is called as "sensory exploitation hypothesis" (Ryan & Keddy-Hector 1992).

In sensory exploitation hypothesis, one of the most important questions to understand the evolution of male traits is to know why the bias arises in female sensory system and the direction of the bias that is likely to appear. A monumental and brilliant paper which answered this question was written by Enquist & Arak (1993). They noted that nerve and sensory systems in animals are often formed through training and learning procedures after they are born. The connection among neurons and the structure of brain are not determined genetically, but they are determined by the experiences of the individual. This might cause a tremendously important bias in the recognition mechanisms of animals, thought by Enquist & Arak (1993, 1994). To show this idea, they study the bias that spontaneously occur in a neural network model thorough a simple training to distinguish two similar objects.

The neural networks can perform recognition and decision-making, and are very simple models for nerve systems in animals. They have connections of artificial neurons and have an ability of learning by modifying the connection weights between neurons. They also have an ability of generalization. Imagine a pigeon, for example, is conditioned with a certain stimuli A. After the training is over, it responds not only to the A but also responds to some signals which is similar to the A. This phenomenon is named "generalization" and has been commonly shown by animals (Guttman & Kalish 1956). After the training, some signals which are not used in the training let the conditioned animal respond more than the training signals. Those signals explain "super normal stimuli" observed among animals -- animals sometimes respond very strongly to exaggerated stimuli that have never shown to them. Based on their study of neural network models, Enquist & Arak (1993) suggested that super normal stimuli could emerge very easily by the simple discrimination training, and that the super normality occurring in this manner can be the driving force for the evolution of male's exaggerate signals. However this famous work by Enquist & Arak (1993) was severely criticized by Kamo *et al.* (1998), who suggest that their results are simply due to their wrong choice of training procedures. Kamo *et al.* demonstrated that no significant supernormality should evolve if the sufficiently long and careful training is

done, showing the care needed in carrying out the study based on the neural network modeling.

The largest advantage of adopting neural network models in evolutionary biology and ecological sciences, instead of more traditional models for games, optimization, and quantitative genetics, is that neural networks can express any complex input-output relationship very easily even when there are multiple input. The evolution of animal signals can be discussed by traditional models only when it can be represented in a single quantity, such as "length of a tail", "loudness of a call", and "size of eye spots". However when we need to study the symmetry of the pattern of body coloration, or the harmony of auditory signals in courtship, these cannot be modeled properly without using neural network models. In this thesis, I study the use of neural network models in evolutionary biology and ecology that can handle the aspects that cannot be modeled by traditional modeling techniques.

In chapter 1, I made a criticism of Enquist & Arak (1993). I trained a neural network which have the same architecture and also the training images were the same in Enquist & Arak (1993) Contrary to the them, I did not observed any bias in the network. Hence I concluded that their result was completely wrong.

In chapter 2, the evolution of the preference for consonances is studied. The combinations of sounds are called as chord. In human, chords are divided into two categories; one is consonances which is comfortable to listen. The other is dissonance which is not comfortable to listen. I studied why we humans show a preference to consonances. I used three-layered neural networks which were a model of auditory system of ears. When the networks were trained with the sounds with harmonics, the network showed the preference for consonances. This is the first research which revealed why we have a preference for consonances.

In chapter 3, I studied the evolution of signal form under both genetically inherited and individually learned recognition. In animals, both genetic and learned recognition occurs in nature. For example, frog females recognize the call of their conspecific males without any learning. On the other hand, some predators learn to

avoid unsuitable prey in their lifetime and the learned ability is not inherited to next generation. I simulated coevolution between senders and receivers using artificial neural networks as the receiver's recognition mechanism.

In chapter 4, I studied the forecasting of temporal processing data using artificial neural networks. In this study, the networks were used as the non-linear model for forecasting the number of infected by echoviruses. The network receives past n time series data and predicts next time step. The network's prediction power was estimated by using the correlation between observed data and the predicted data. In the prediction of the number of hosts infected by echoviruse, the correlation was quite high (around 0.8).

In the following, I explain the contents of these four chapters in more detail:

Chapter 1. Neural Network for Female Mate Preference, Trained by a Genetic Algorithm

In some animals, males evolve exaggerated traits (e.g. the peacock's conspicuous tail and display), due to female preference. Recently Enquist & Arak presented a simple neural network model for a visual system in female birds that acquires the ability to discriminate males of the correct species from those of the wrong species by training. They reported that the trained networks were attracted by "supernormal stimuli" where there was a greater response to an exaggerated form than to the images used as the correct species for training. They suggested that signal recognition mechanisms have an inevitable bias in response, which in turn causes selection on signal form. I here examine the Enquist & Arak model in detail. A three-layered neural network is used to represent the female's mate preference, which consists of 6x6 receptor cells arranged on a regular square lattice, 10 hidden cells, and one output cell. Connection weights of the network were modified by a genetic algorithm, in which the female's fitness increases if she accepts a conspecific male but decreases if she accepts a male of a different species or a random image. I found that: [1] After the training period, the evolved network was able to discriminate male images. Female preference evolves to favor unfamiliar

patterns if they are similar to the images of the correct species (generalization). [2] The speed and the final degree of learning depended critically on the choice of the random images that are rejected. The learning was much less successful if the random images were changed every generation than if 20 random images were fixed throughout the training period. [3] The male of the same species used for training achieved the highest probability of being accepted by the trained network. Hence, contrary to Enquist & Arak, the evolved network was not attracted by supernormal stimuli.

Chapter 2. Evolution of Preference for Consonances as a By-product

Recent theoretical study of evolution of visual signals in animals has revealed that biased preferences for symmetric patterns or simple coloration can evolve in the absence of positive fitness effects. In this paper I study the evolution of biased preference for auditory signals. In music theory, intervals between a pair of auditory signals are classified into consonances and dissonances. Consonances are more comfortable to listen than dissonances, and often have a frequency ratio close to a ratio of small integers. By examining the preferences shown by a three-layered network as a simplified model of an auditory system, I study why we feel consonances comfortable and dissonances uncomfortable. When the network was trained to accept monotones accompanied by harmonic tones and to reject random signals (noises), it developed preference for consonances over dissonances. This suggests that the preference for consonances may have evolved as a by-product of training for a simple task, such as distinguishing mother's voices from noises, rather than as a result of being taught one-by-one. When a network was trained to favor a consonance and to reject a dissonance, it did not generalize the preference to other consonances or dissonances.

Chapter 3. Evolution of Signal Form: The Effects of Individually Learned versus Genetically Inherited Recognition

The evolution of the signal recently has much interest. Most studies assume that the recognition which are transmitted to successive generations and the effect of the

individual learning to the evolution of signal has been ignored. In this study, I focused on the effect of individual leaning. I used three layered neural networks with 10 input cells, 5 hidden cells and an output cell and the cells between layers are fully connected with some weights. In the former studies (Enquist & Arak 1993, 1994, Kamo *et al.*), they used the genetic algorithm in the network training, i.e. both the network recognition and the signal forms were transmitted to the successive generation. To examine the effect of the individual learning, the process of individual learning was modeled as follows: (1) A naive network learns to prefer a signal in a generation for a certain period. (2) After the network's learning is stopped, the mutant signal is created and both the original signal and mutant are shown to the network. (3) A better signal is inherited to the next generation, but the network is not. (4) The weight of the network in new generation is initialized and it starts to learn to prefer the signal. Comparing the results from both the genetically inherited model and individual learning model, I found that more exaggerated signal had evolved in individual learning model than in the genetically inherited model.

Chapter 4. A Neural Network Forecasting of the Viral Subtype Dynamics: Applications to the Echovirus epidemics.

I study artificial neural networks to forecast the outbreaks of infectious disease, which is then applied to echovirus epidemics data. The network's prediction power is tested by time series data generated by the SIR compartment model with seasonally varying transmission rate. The SIR dynamics with seasonality is known to show a cascade of bifurcations as the strength of seasonality increases. I first investigated the bifurcation diagrams of the single and two strain SIR models. The dynamical behavior critically depends on the strength of seasonality, and the degree of cross-immunity in two strain case. I have chosen different parameter sets from each region of the phase diagram, and generated the time series of the SIR model. The neural network trained by the earlier half of the time series was due to predict the later half of the data. I found that when the dynamics shows simple one-year cycles or two-year cycles, the network's forecasting is

very good for quite long time: the correlation between generated data and the predicted data stayed near 1 for long time. When the dynamics shows a chaotic behavior, the network's short-term forecasting is still good (retain high correlation) but the accuracy of prediction of gradually falls off after a few years. Lastly, the network was trained with the reported data for the echoviruses outbreaks during 1983-1997 in Japan. The correlation between the observed data and that predicted one by the network was as high as 0.8 even at 7 years after the starting point of forecasting. The result shows that neural network could work as an efficient statistical model for the forecasting of the epidemical time series.

In spite of the large merit of introducing neural network models that can handle many diverse problems in evolution and ecology that cannot be handled otherwise, they have limitations. In search for the optimal solution, we must use random search techniques, such as genetic algorithm (GA), back propagation, self-organization mapping. At this moment we do not have systematic study of the choice of architecture of networks and important parameters for training procedures, such as mutation and crossing-over rates and the population size in GA. In engineering, these are done simply in tries and errors. As a consequence a large mistake in biological conclusion can be made as shown by Enquist & Arak (1993) (see Kamo *et al.* 1998). To avoid this, we need to apply multiple methods to the same model and also have to check the robustness of the conclusion by modifying the architecture of the networks and the training algorithm. However I am very positive in the future of these techniques. The speed and the memory size of computers become exponentially increasing every year, and these problem can be overcome if we carefully check the robustness of the conclusion, and the same model can be tested by separate group of people immediately. In the near future, these computation-heavy modeling will become to be the mainstream of the theoretical modeling in ecology and evolutionary biology.

References

Enquist, M. & Arak, A. 1993. Selection of exaggerated male traits by female aesthetic senses. *Nature, Lond.* **361**, 446-448.

Enquist, M & Arak, A. 1994. Symmetry, beauty and evolution. *Nature, Lond.* **372**: 169-172.

Fisher, R. A. 1930. *The genetical theory of natural selection.* Oxford: Clarendon Press,

Grafen, A. 1990. Biological signals as handicaps. *J. Theor. Biol.* **144**:517-546.

Guttman, N. & Kalish, H. I. 1956. Discriminability and stimulus generalization. *J. Exp. Psychol.* **51**: 79-88.

Iwasa, Y., Pomiankowski, A. & Nee, S. 1991. The evolution of costly mate preferences. II. The 'handicap' principle. *Evolution* **45**, 1431-1442.

Kamo, M., Kubo, T. and Iwasa, Y. 1998. Neural network for female mate preference, trained by a genetic algorithm. *Phil. Trans. R. Soc. Lond. B*, **353**: 399-406.

Pomiankowski, A., Iwasa, Y. & Nee, S. 1991. The evolution of costly mate preferences. I. Fisher and biased mutation. *Evolution* **45**, 1422-1430.

Ryan, M. J. & Keddy-Hector, A. 1992. Directional patterns of female mate choice and the role of sensory biases. *Am. Nat.* **139**, S4-S35.

Zahavi, A. 1977. The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603-605.

CHAPTER 1

Neural Network for Female Mate Preference, Trained by a Genetic Algorithm

This chapter was done in collaboration with Dr Takuya Kubo and Professor Yoh Iwasa. The paper was published in Philosophical Transaction Royal Society of London Series B, (353: 399-460) in 1998.

INTRODUCTION

Males of some animals have exaggerated traits, which are apparently disadvantageous to the holder's survivorship. A classical example is the tail of the male peacock. Darwin (1871) suggested that males develop a structure costly to produce and maintain because females have a propensity to choose mates with such exaggerated traits. After Darwin, numerous experimental and theoretical studies have been carried out on female mate preference (reviewed by Andersson, 1994; Moller, 1994; Andersson & Iwasa, 1996).

Fisher (1930) was the first to succeed in explaining the evolution of female preference for an exaggerated male trait. If females in the initial population have a small propensity to choose mates with a longer tail, then a female with a stronger preference than the population average can enjoy an indirect benefit of producing sexy sons who inherit their father's "good looking" long tail. As a consequence, exaggerated male traits and strong female preference will coevolve simultaneously, resulting in Fisherian runaway selection (Fisher, 1930; Lande, 1981; Pomiankowski *et al.*, 1991; Pomiankowski & Iwasa, 1993; Iwasa & Pomiankowski, 1995). According to this argument, any male trait and a female preference for it can coevolve.

In contrast, Zahavi (1977) noted that male traits used for sexual selection are often good indicators of male quality, revealing physical strength, mental ability, and resistance to parasites, traits that are difficult to determine directly. A visible male trait can be used by females in their mate choice, when only a really strong male can afford to produce and maintain it. The peacock's tail is a trustworthy signal of a good and strong male simply because it is a costly handicap to the holder (Hamilton & Zuk, 1982; Grafen, 1990; Iwasa *et al.*, 1991; Iwasa & Pomiankowski, 1994).

A third theory emphasizes the inherent bias of the sensory organ (Kirkpatrick & Ryan, 1991; Ryan & Keddy-Hector, 1992). If females are attracted by stronger stimuli, such as a louder call, or a longer tail, then males may evolve to exploit that sensory bias by exaggerating signals.

Why is there bias in sensory organs? Enquist & Arak (1993) and Arak & Enquist

(1993) developed a simple neural network model to imitate a signal recognition system that is adjustable by training, and demonstrated that the trained network has an inherent bias. The neural networks had an artificial retina on which simple bird-like images were presented. They were trained to discriminate two simplified images: one represents a conspecific male bird with a long tail (figure 1c), and the other pattern with a short tail for the male of a wrong species (figure 1b). After the networks were trained to discriminate between them, some new images were presented. The network responded to some of the novel images that are similar to the stimuli used in training, which implies "generalization" (Guttman & Kalish, 1956; Hanson, 1959). In some case, a new stimulus gave a response even stronger than the image of the correct species used in the training, and hence it is a "supernormal stimulus" (Staddon, 1975). These images had a tail longer than the correct stimulus, and were thus an exaggerated form (figure 1d). The reaction for signals of different tail length had a peak shifted from the one that had been used in training. Enquist & Arak suggested that supernormal stimuli and "peak shift" are a natural outcome of the sensory mechanism formed by training, and that a biased signal recognition system might be the basic mechanism for the evolution of exaggerated male secondary traits.

In this paper, we examine Enquist & Arak's (1993) system carefully and attempted to confirm their results by training a neural network by genetic algorithm. After training, the network was able to distinguish males of the correct species and wrong species. In addition, the trained neural networks show the greatest response to the pattern used for training, implying that there was no supernormal stimuli. We will discuss possible reasons for the discrepancy between Enquist & Arak's and our study.

NEURAL NETWORK MODEL

Neural networks are mathematical models imitating the function of a network of neurons (neural cells) in a computer. Neural network can learn to distinguish patterns. Each neuron, the unit of a neural network, receives n input signals $x_1, x_2, ..., x_n$, with weights $w_1, w_2, ..., w_n$ that may differ between inputs. A negative weight (w_i<0) implies that the corresponding input has an inhibitory effect. The simplest deterministic version of the model assumes that the neuron responds by outputting 1 when the weighted sum of the input signals is larger than a threshold denoted by h. If not, it outputs nothing. Due to noises from various sources, a neuron is likely to respond stochastically with a probability given by a sigmoid function of the sum of the input signals. We assume that the probability of getting the reaction 1 from the neuron is a sigmoid function:

$$\begin{bmatrix}\text{Probability of} \\ \text{output = 1}\end{bmatrix} = \frac{1}{1+\exp\left[-a\left(\sum_{i=1}^{n} w_i x_i - h\right)\right]} \quad . \qquad (1)$$

Figure 2 is the neural network used in Enquist & Arak (1993) and in the present paper. It consists of an artificial retina of 6x6 receptor cells, 10 cells in the hidden layer, and one output cell. Each cell in one layer is connected to every cell in the next layer, each connection being accompanied by a weight. The learning progresses by changing the weights appropriately. Male bird-like patterns were projected onto the retina (receptor cell layer), and a male is accepted by the female network as a mate when a positive response from the output cell is given.

The images used for training and testing of the networks are shown in figure 1 as filled (black) squares on a 6_6 retina. The bird-like image figure 1c with tail length two was defined as a male of the same species, and the image of figure 1b with tail length one was considered as a male of a different or wrong species. The correct response of the female is to accept figure 1c and to reject figure 1b.

Random images were also used as reject images. Enquist & Arak did not describe clearly how to generate the random images they used. We made random images with 5 to 8 filled squares (figure 3). We chose the number of filled squares with care in order to avoid the situation in which the total number of filled squares can be used as a useful cue to distinguish correct and wrong species.

When presented to the networks for training, both the patterns of tail length two (the correct species) and tail length one (wrong species) were moved and rotated on the retina, but the 20 random images were fixed. Considering the number of different positions and orientations, there are 48 possible configurations of tail length two, and 16 possible patterns of tail length one.

(a) *Genetic algorithm learning*

Back propagation is the learning method most frequently adopted when correct reactions (teach signals) of the network are given (Welstead, 1994). Weights are appropriately changed to decrease the difference between the output by the network and the teach signal. However we adopted an alternative learning method, called genetic algorithm (Davis, 1990; Michalewicz, 1994), that was used in the modeling evolution of signaling (Arak & Enquist , 1993; Enquist & Arak, 1993; Johnstone, 1994). Recently, a neural network model learned by both genetic algorithm and back propagation was applied to a study of the artificial life (Toquenag *et al.* 1995).

We consider a string of connection weights of a neural network as one full set of genes (or genome), which can specify an individual network. We then consider a population of 20 individuals (or 20 strings of connection weights). Selective reproduction and mutation of these individuals allow the evolution of the network to make the neural network discriminate the images. The procedures are:

(1) initial configuration

The initial values of the connection weights were generated by random numbers uniformly distributed between -1 and 1.

(2) calculating scores

In every generation, various images were presented to each of the 20 networks in

the population. Each network (or individual) experienced 150 input images per generation. Fifty of these images were correct species, which are chosen randomly (with resampling allowed) from 48 possible configurations, considering shift and rotation. Another fifty images are on the wrong species, which are chosen randomly from 16 possible configurations. Then the remaining fifty were randomly sampled from 20 possible random images in figure 3.

When the network responded (i.e. the output was 1) to the image of the correct species, it gained one score. If instead the network responded to the image of the wrong species or to random images, then its score was reduced by one.

(3) reproduction

When each of 20 networks in the population had experienced exposure to 150 images, we calculated the accumulated score of each network. The ten networks with the highest score were allowed to have two copies of themselves in the next generation. Ten networks with the lowest scores had no possibility of contributing to the next generation. In this way, the population size was kept to 20 all the time. In terms of population genetics, we adopted a "truncation selection" scheme with respect to the score.

(4) mutation

Mutation then occurred on the connection weights. The probability of mutation for a particular weight was one tenth and when mutation occurred, an increment drawn from a normal distribution (mean = 0, standard deviation = 0.4) was added to the weight.

These steps were iterated 10,000 generations (training period). The networks converged to some kinds of equilibrium by then, and the average responses to learning tasks did not change so much even if the training period was doubled. The iteration of these procedures could develop a network that can discriminate males of the same and different species.

(b) *Test period*

After the training period, we presented various images, to the educated networks

without modifying the weights, and examined their reaction to the test images. In the test period, the connection weights were fixed. The test images included the ones used in the training, but also those that have never been shown to the networks, as illustrated by (a), (d) - (g) in figure 1.

To count the strength of reaction of a network to an image, we chose 100 images randomly among all the patterns generated by the shifting and rotation of the original image. We then calculated the "average response", the number of times in which the network gave output 1 among 100 trials. To examine the reaction of a particular pattern to "random images", we examined 100 random samples (with resampling allowed) from the same 20 random reject images fixed throughout the training period.

In addition to these standard training procedures, we also examined the several different training schedules, as explained below.

RESULTS

Results of the training experiments were as follows:

(a) *Average response*

In the initial population of 20 networks of randomly generated connection weights, the average response differed greatly between networks, but were similar between different images. Some networks responded to most of the input images, while others rejected most of the images. After the training period, the average response became similar between networks in the population and different between images.

Figure 4A illustrates the average response, of 100 trials for each of the 20 networks in the final population. The average response for the pattern with tail length two (correct species) was quite high (98.1%), and that for the pattern with tail length one (wrong species) was very small (2.6%).

(b) *Absence of supernormal stimuli*

In the test period, various images other than those used for training were presented to the networks (figure 1), but no image had a higher average response than the images of the tail length two (figure 4A). The average response to a pattern with a longer tail (tail length three) than the conspecific male (tail length two) was quite high, but was lower than that of the conspecific male (Wilcoxon signed rank test, $z = -3.926$, $p < 0.0001$). We did not therefore observe supernormal stimuli. This is very different from the result of Enquist & Arak, in which the average response was always higher for an image with a longer tail.

Several replicates were calculated for the same set of parameters. The results described above held in all the cases. The connection weights of the networks obtained in the final population were totally different, but they behaved in a very similar way.

Since we examined only patterns with tails of integer length (e.g. tail length 1, 2, and 3), we may not be able to detect a peak shift of a small magnitude even if it exists. To overcome this difficulty, we examined responses of the trained network to patterns with a tail of noninteger length. For example, a pattern of tail length 1.75 implies a

pattern with tail length 1 appended by a "grey" unit with intensity of 0.75. Figure 4B illustrated the results, which shows the pattern with tail length 2 (correct species) caused the highest average response of the network, and hence there was neither peak shift nor supernormality.

(c) *Trajectory of learning procedures.*

Figure 5 illustrates the time course of the average response of a population to images of three classes: the images of correct species, those of wrong species, and random images. In every generation a hundred images from each of these three classes were presented to the networks, and the average responses in figure 5 are the proportion of times at which the network responded positively. The line with symbol II is for the response to the images of correct species (tail length two), which increases with time and reached a level of almost 100% at the end of the training period. The line with symbol I is for the average response of the network to the images of wrong species (tail length one), which became very low and close to zero. The reaction to random reject images with symbol R also became lower with time, though it fluctuates greatly.

The broken line with symbol III shows the response by the network to the images of tail length three (with shift and rotation). The responses of the network to this class of images were just measured but not used for training the network (unlike the responses indicated by the other three lines).

The response to the image of tail length three wildly fluctuated with time but was consistently lower than the response to the image of tail length two (correct species). Hence there was no indication that supernormal stimuli might cause a stronger response than the stimuli used for training.

(d) *Different choice of random images*

In the standard procedures of training, we fixed 20 random images and used them as reject images throughout the training period. Alternatively, we may generate different random images in every generation. Figure 6A illustrates the average reaction of the network trained with such random reject images. The average reaction for the image of the conspecific males was 100 %. However the average response for the image

of the wrong species was not small, remaining at around 20%, even after 10,000 generations. The average reaction to random images remained at 60%. Hence we can conclude that the learning was not as effective if the random images changed in every generation.

Second, in the standard training procedures, we used 20 random images that include images of 5 to 8 filled squares. This is so that the network cannot use the total number of black squares for the purpose of distinguishing the images.

We have also trained using 20 random images, which all have 5 filled squares. The results in figure 6B do not show clear differences from the standard training procedure of random images of 5-8 filled squares (figure 4A). However, if random reject images are of 2 filled squares only, the trained networks show a similar but smaller response to some images with larger number of filled squares (figure 6C). For example the difference between (c) and (f) in figure 6C was not significant (Wilcoxon signed rank test, $z = -0.784$, $p = 0.4331$).

Third, we have done training the networks without random images. The results are shown in figure 6D. The average response to the correct species was again significantly higher than that to the other images (Wilcoxon signed rank tests with Bonferroni adjustment, $|z| > 3.922$, $p < 0.0026$). Interestingly the average response to the wrong species (b) was very low but the responses to unfamiliar patterns were higher than the case random reject images were used. We suspect that this enhanced acceptance rate might be related to the reduced total number of reject images used during training.

(e) *Different number of images to present per generation*

In the standard training procedure, each network experiences 150 images every generation. We also examined cases with fewer images per generation for training. In a particular case, for example, we presented all images once per generation, so that 48 images of the correct species (including shift and rotation), 16 images of the wrong species (including shift and rotation), and 20 images of random reject images. The learning progressed somewhat slower than the standard training procedures.

Interestingly, networks that accepted all the patterns increased and dominated the population temporarily but then slowly disappeared. The final outcome of the evolution was almost the same as the standard training procedures.

DISCUSSION

There was a major difference between the network trained in the present paper and the one reported in Enquist & Arak (1993): We did not observe supernormal stimuli, as illustrated in figure 4A and 4B.

Since the network structure is exactly the same in Enquist & Arak's study and ours, the difference must be due to the training procedures. From the comparison, we must conclude that Enquist & Arak's (1993) networks might have been trained less than perfectly. However, Enquist & Arak did not describe in detail the procedures they used for training the network. In the following we would like to list potential processes that might have caused the difference in the effectiveness of the learning processes, irrespective of their applicability to Enquist & Arak's training procedures.

First and the simplest possibility is that the they stopped training before the network achieved the full ability to distinguish the patterns.

The second possibility is that they might have used a selection scheme that was less efficient than the truncated selection adopted in our training scheme. If we used a genetic algorithm with the fitness proportional to the contribution to the following generation, the training would be slower than in the case in which ten networks with the higher score remain and the other ten of lower score are discarded.

Third, the choice of random reject patterns used during training requires some care. In our training scheme, when we made random images, we chose the number of filled squares so that the networks could not distinguish images based on the total number of filled squares. If we used random images including only a few filled squares, images with two wings longer than the correct species were accepted by the network nearly as often as the correct species (figure 6C). Training without using random images at all showed no indication of supernormality but generally high response to unfamiliar images (figure 6D).

Fourth, we fixed 20 random images which we used as reject images throughout the training and test periods (figure 3). If instead we used random images generated

each generation, the learning efficiency was not very high (figure 6A).

Fifth there are other relatively minor differences between Enquist & Arak and our networks. For example, the sigmoid curve for each neuron used by E & A was an integral of a normal distribution, which is different from Eq. (1) used in our training, the latter being more commonly used in neural network models (Welstead, 1994). However we do not expect this difference is the major reason of the difference in efficiency in training.

Sixth, a possible reason of the absence of supernormality is that the neural network might have been over-trained. Over-training would reduce the generalization ability lower. To examine this possibility, we run simulation in which four input patterns to accept (among 48 configurations considering shift and rotation) and one reject pattern (among 16 configurations) were not shown to the network during training. We then examined the average response of the networks at different times (1000, 2000, 3000, 4000, and 10000 generations). The reaction to these images excluded in training fluctuate considerably, but the trend was clear: as training proceeds, the networks became to accept correct stimuli more often, hence generalization tendency increased, rather than decreased. On the other hand they also accept one image of tail length 1 (wrong species) throughout the training period. In short, there was no indication of overtraining because the tendency of generalization did not decline with time.

Generalization

In the present study, the average response to the image of tail length three was higher than other images except the one used for training (tail length two) (Wilcoxon signed rank tests with Bonferroni adjustment, , $|z| > 3.922$, $p < 0.0004$). The image of tail length three was accepted because it resembles the conspecific male image. This illustrates that recognition by a trained neural network also shows generalization.

Generalization is a common property of recognition systems which classify unfamiliar stimuli into familiar categories. A classical example of generalization is that of a pigeon trained to perform a behavior if it observed a light bulb of wave length 550

Mμ (Guttman & Kalish, 1956; Hanson, 1959). Naturally the response of the pigeon was the highest when stimulated by the same color. However the pigeon also responded to a light of similar wave length.

The results we obtained for the present study can be understood intuitively in the following way (figure 7). The positive training by images of the correct species results in the network favoring patterns similar them, illustrated by a broken line in figure 7. The resulting reaction of the trained network has some breadth around the peak reaction corresponding to the image of the correct species. The curve is symmetric around the peak, and hence the generalization is not biased. On the other hand, negative training by images of wrong species males would make the network to avoid patterns similar to the image of wrong species, as indicated by another line in figure 7. A combination of these two tendencies would be simply the sum of the two curves, resulting in the observed pattern of the average responses to the tail images of different tail length (figure 7). Spence (1937) proposed a similar idea : a combination of positive and negative generalizations might explain an animal's response after training.

If this simplified picture holds, we may be able to infer the generalization pattern when choices of male images for correct and wrong species are different from the one adopted in the present paper. For example, suppose the images of the tail length three are defined as conspecific species males and those of tail length one are defined the wrong specie males. Probably, the networks can discriminate these two classes of images almost perfectly and the average response to the images of tail length four will be high but probably not higher than the reaction to conspecific pattern (tail length three). Average response to the pattern of tail length two will be intermediate because it is affected by both the positive generalizations of the conspecific pattern of tail length four and the negative generalization of tail length one. Although the distances from the optimum (tail length three) of tail length two and tail length four are the same, the acceptance rate was different . In this case, sensory bias could be more apparent. Whether the scheme in figure 7 can work in this or other situations will be an important subject of future study.

Whatever the reasons, we did not see supernormality in spite of our effort to reproduce the same procedures described in Enquist & Arak (1993), in which unfortunately the computational procedures were not explained well. In the present paper, we tried to present our work in sufficient detail that any interested reader can redo our training procedures. The only case in which we have closest to the reported in Enquist & Arak may be the case in which the network was trained with additional random reject patterns of very few black squares, but even then the average reaction of an images with two longer wings ((f) in figure 6C) was not higher than the correct images ((c) in figure 6C). We must conclude that we have not obtained a convincing case of supernormality for various modifications.

Inspired by Enquist & Arak's pioneering work, there have been developed numerous theoretical studies on the property of sensory systems that are adjusted by training procedures (Johnstone, 1994; Enquist & Arak, 1994; Arak & Enquist, 1995). Using the same models as Enquist & Arak (1993), Arak & Enquist (1993) trained networks to discriminate flowers that had petals of different lengths, which resulted in a bias in pollinators' preference. The evolution of symmetrical visual patterns was discussed in Enquist & Arak (1994) and Johnstone (1994). Although these works have received some criticisms as being too simplistic (Cook, 1995; Dawkins & Guilford, 1995), we believe that modeling preference and choice systems as a simple neural network that can be trained through experiences is a very promising approach in the study of biological signals. However, to derive robust and useful conclusions, it is important to examine the training procedures carefully, as we have done in this paper.

REFERENCES

Andersson, M. B. 1994 *Sexual selection.* New Jersey: Princeton University Press.

Andersson, M. B. & Iwasa, Y. 1996 Sexual selection. *TREE* **11**, 53-58.

Arak, A & Enquist, M. 1993 Hidden preferences and the evolution of signals. *Phil. Trans. R. Soc. Lond.* B. **340**, 207-213

Arak, A. & Enquist, M. 1995 Conflict, receiver bias and the evolution of signal form. *Phil. Trans. R. Soc. Lond.* B. **349**, 337-344

Cook, N. D. 1995 Artifact or network evolution? *Nature, Lond.* **374**, 313-314.

Davis, L. 1990 *Hand book of genetic algorithms*. New York: Van Nostrand Reinhold.

Dawkins, M. S. & Guilford, T. 1995 An exaggerated preference for simple neural network model of signal evolution? *Proc. R. Soc. Lond.* B. **261**, 357-360.

Enquist, M. & Arak, A. 1993 Selection of exaggerated male traits by female aesthetic senses. *Nature, Lond.* **361**, 446-448.

Enquist, M & Arak, A. 1994 Symmetry, beauty and evolution. *Nature, Lond.* **372**: 169-172.

Fisher, R. A. 1930 *The genetical theory of natural selection.* Oxford: Clarendon Press,

Grafen, A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**:517-546.

Guttman, N. & Kalish, H. I. 1956 Discriminability and stimulus generalization. *J. Exp. Psychol.* **51**: 79-88.

Hamilton, W. D. & Zuk, M. 1982 Heritable true fitness and bright birds: a role for parasites? *Science* **218**, 384-387.

Hanson, H. M. 1959 Effects of discrimination training on stimulus generalization. *J. Exp. Psychol.* **58**, 321-334.

Iwasa, Y. & Pomiankowski, A. 1994 Evolution of mate preferences for multiple sexual ornaments. *Evolution* **48**, 853-867.

Iwasa, Y. & A. Pomiankowski. 1995 Continual change in mate preferences. *Nature* **377**: 420-422.

Iwasa, Y., Pomiankowski, A. & Nee, S. 1991 The evolution of costly mate preferences. II. The 'handicap' principle. *Evolution* **45**, 1431-1442.

Johnstone, R. A. 1994 Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature, Lond.* **372**, 172-175

Kirkpatrick, M. & Ryan, M. J. 1991 The evolution of mating preferences and the paradox of the lek. *Nature, Lond.*. **350**, 33-38.

Lande, R. 1981 Models of speciation by sexual selection of polygenic traits. *Proc. Nat. Acad. Sci.* USA **78**, 3721-3725.

Michalewicz, Z. 1994 *Genetic Algorithms + Data Structures = Evolution Programs*. (2nd ed.) New York: Springer-Verlag.

Moller, A. P. 1994 *Sexual selection and barn swallow*. New York: Oxford University Press

Pomiankowski, A. & Y. Iwasa. 1993 Evolution of multiple sexual ornaments by Fisher's process of sexual selection. *Proc. Roy. Soc. Lond. B.* **253**: 173-181.

Pomiankowski, A., Iwasa, Y. & Nee, S. 1991 The evolution of costly mate preferences. I. Fisher and biased mutation. *Evolution* **45**, 1422-1430.

Ryan, M. J. & Keddy-Hector, A. 1992 Directional patterns of female mate choice and the role of sensory biases. *Am. Nat.* **139**, S4-S35.

Spence, K.W. 1937 The differential response in animals to stimuli varying within a single dimension. *Phychol. Rev.* **44**:430-444.

Staddon, J. E. R. 1975 A note on the evolutionary significance of "supernormal" stimuli. *Am. Nat.* **109**, 541-545.

Toquenag, Y., Kajitani, I. & Hoshino, T. 1996 Egrets of a Feather Flock Together *Artificial Life*. **1**, 391-411.

Welstead, T. 1994 *Neural network and fuzzy logic application in C/C++*. New York: John Willey & Sons.

Zahavi, A. 1977 The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603-605.

Figure Legends

Figure 1. Images used during the training period and the test period. During the training period, the image (c) of tail length two was used as a right male, and the image (b) of tail length one was used as a wrong male, and they were presented to the network with shifting and rotations, together with 20 random images. During the test period, various images illustrated in this figure were presented to their network without changing connecting weights.

Figure 2. The structure of the network. Input image was regarded as a male's pattern. A female's recognition system is modeled as a three-layered neural network. The input layer includes 6_6 receptor cells arranged on 2-dimensional square lattice, which are connected to 10 cells in the hidden layer, and then to a single output cell. Each cell follows stochastic rule given by Eq. (1), where connection weights are modified during the training period (modified from Enquist & Arak, 1993).

Figure 3. Twenty random images used for training. These 20 images are reject images used in training period and test period. The number of filled squares ranged from 5 to 8. If we generate different random reject images every generation, instead of a fixed set, the speed of the learning of the network was much slower.

Figure 4. The average response of the trained network. (A) The acceptance rate was low (2.6%) for a wrong species image (b), and high (98.1%) for a right species image (a). (B) The reaction of the trained network to patterns with intermediate tail length. To express patterns with a noninteger valued tail length, a pattern of tail length 1.75 implies a pattern with tail length 1 appended by a "grey" unit with intensity of 0.75. Horizontal axis is the tail length (1, 1.25, 1.5, .., 3). The pattern with tail length 2 (used as the correct signal in training) caused the highest average response of the network, and

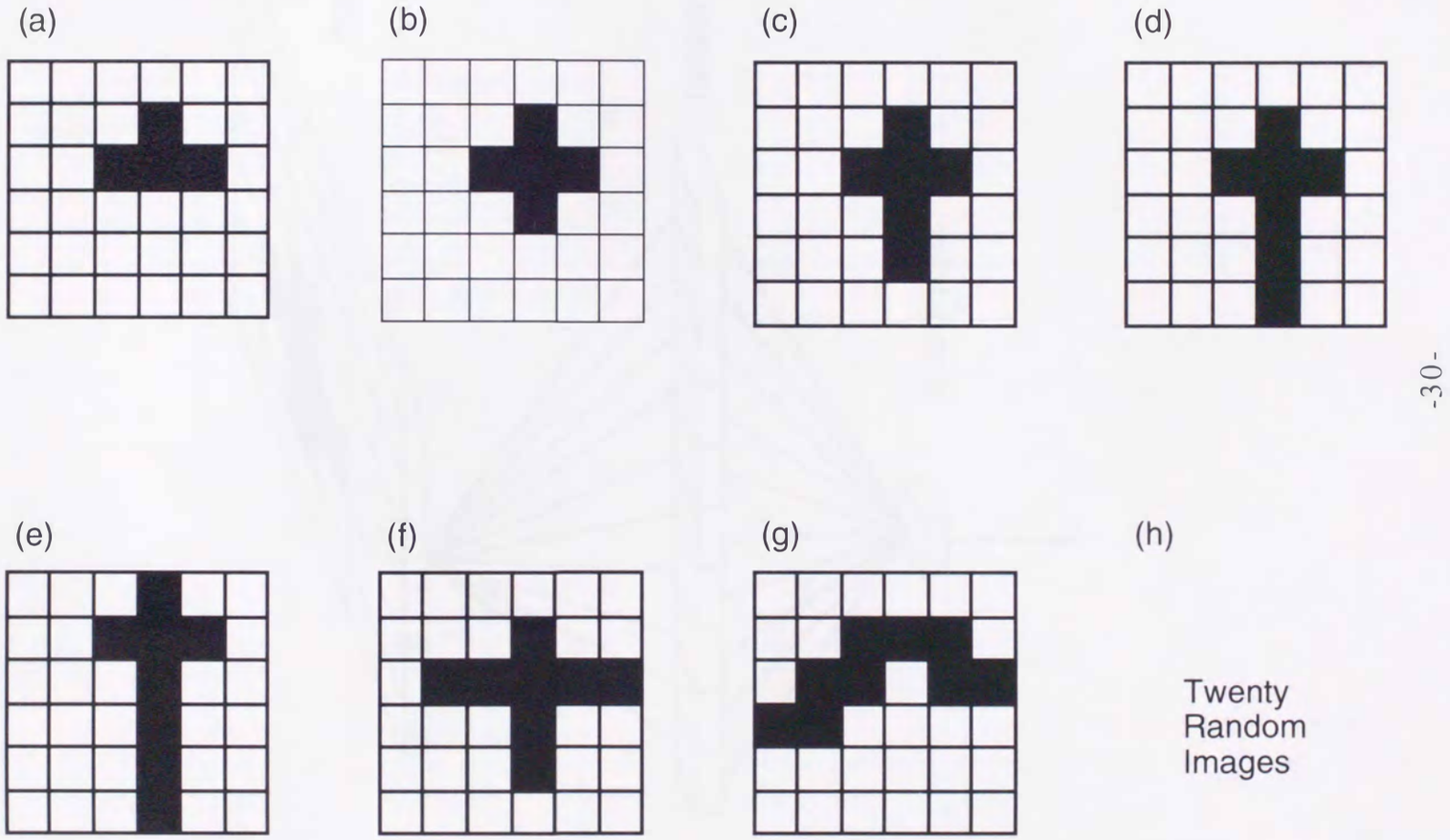
hence there was neither peak shift nor supernormality.

Figure 5. The average response changing with time during training period. Four lines are the response for the images of the correct species with tail length two (with symbol II), the image of a wrong species with tail length one (I), and random images (R). The average response is the fraction of times when the network output 1 among given classes of images. Connection weights were modified by genetic algorithm. The curve of a broken line (with symbol III) is the average response to the image of tail length three. The response to this class of images was not used for training.

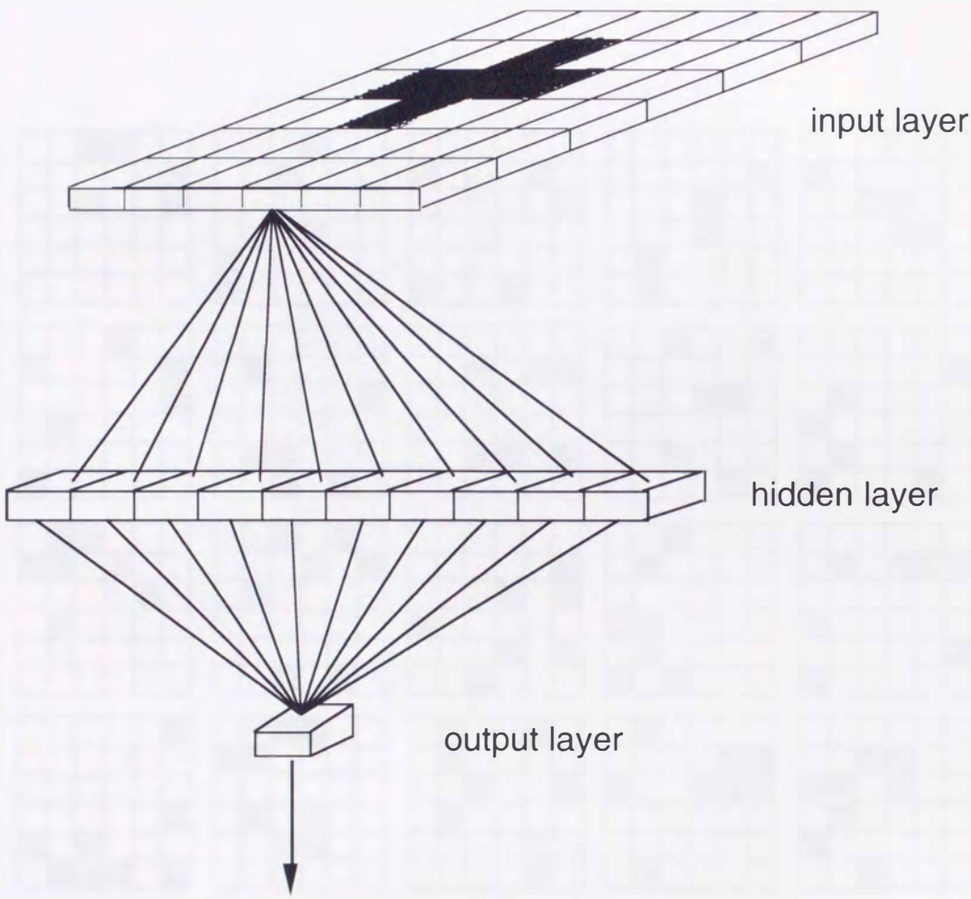
Figure 6. The average response of the trained network when different training schemes were adopted. (A) Different random images were chosen each generation. (B) All the random reject patterns have 5 filled squares, but are fixed throughout the training period. (C) All the random reject patterns have 2 filled squares, but are fixed throughout the training period. The notations and symbols are the same as in figure 4A. The pattern with two longer wings (f) caused a similar response to the conspecific (c). (D) Training without using random reject pattern at all. The trained network tends to accept unfamiliar patterns at a higher rate than in the standard training procedure. These figures demonstrate the importance of the choice of random reject patterns.

Figure 7. A scheme of generalization, summarizing the result of the learning simulation in the present paper.

Chapter 1. Figure 1. M. Kamo

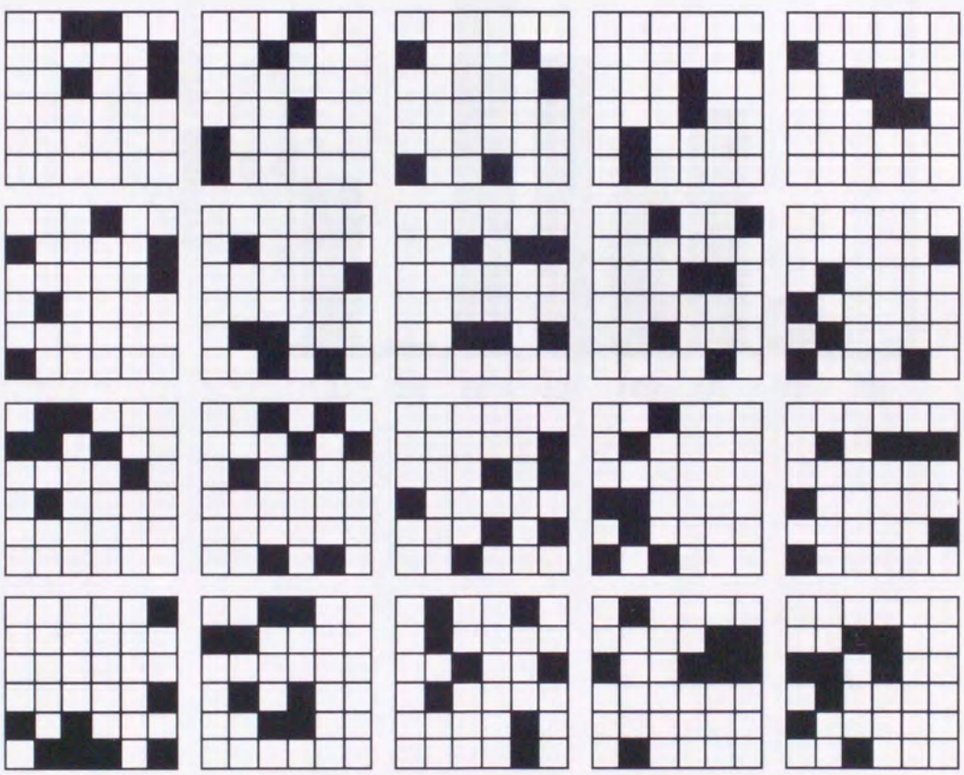


Chapter 1. Figure 2. M. Kamo



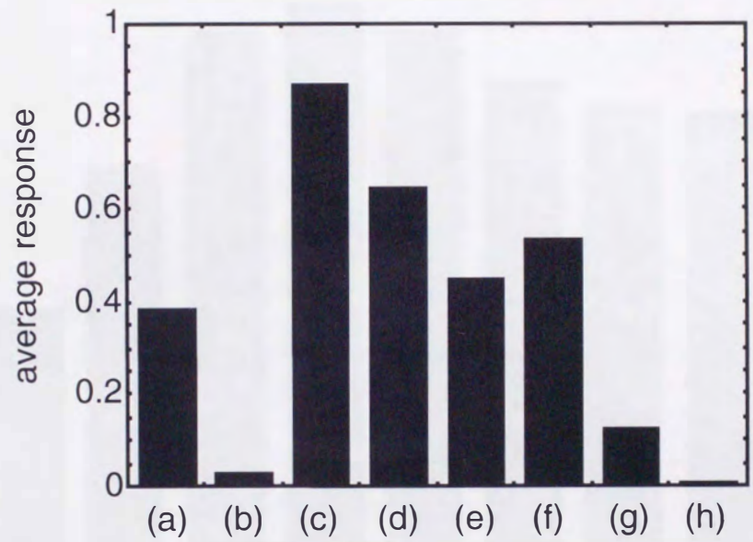
Chapter 1. Figure 3. Kamo

random images with 5- 8 dots

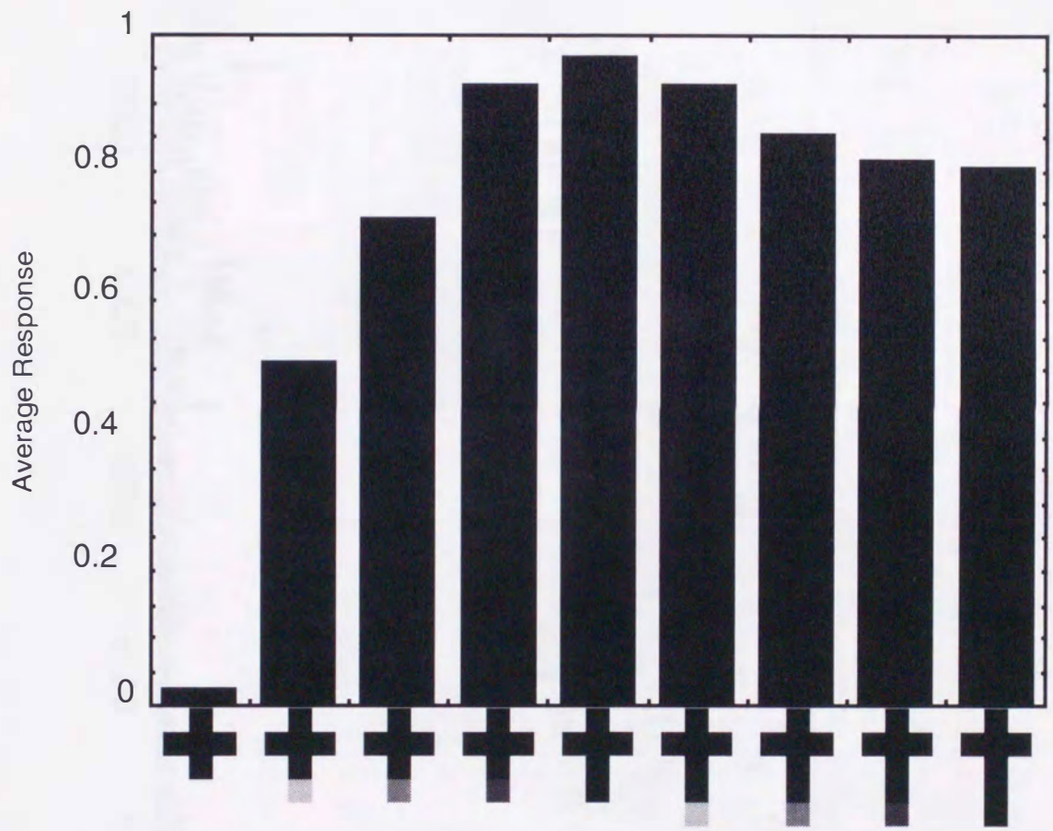


Chapter 1. Figure 4A. M. kamo

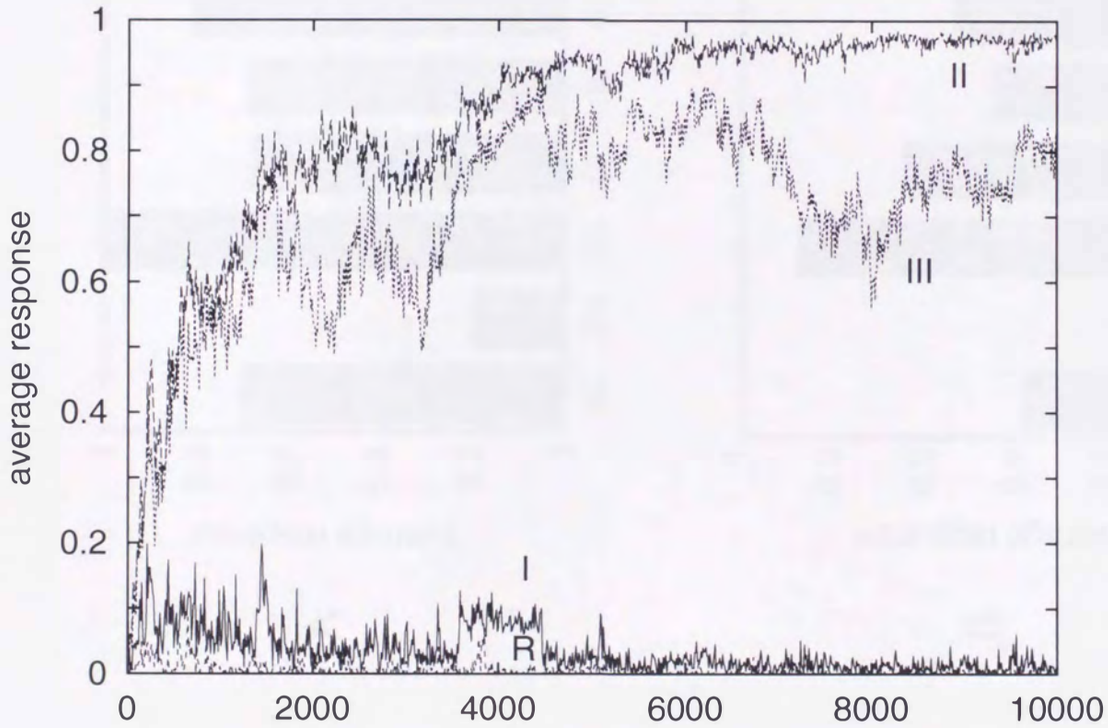
A



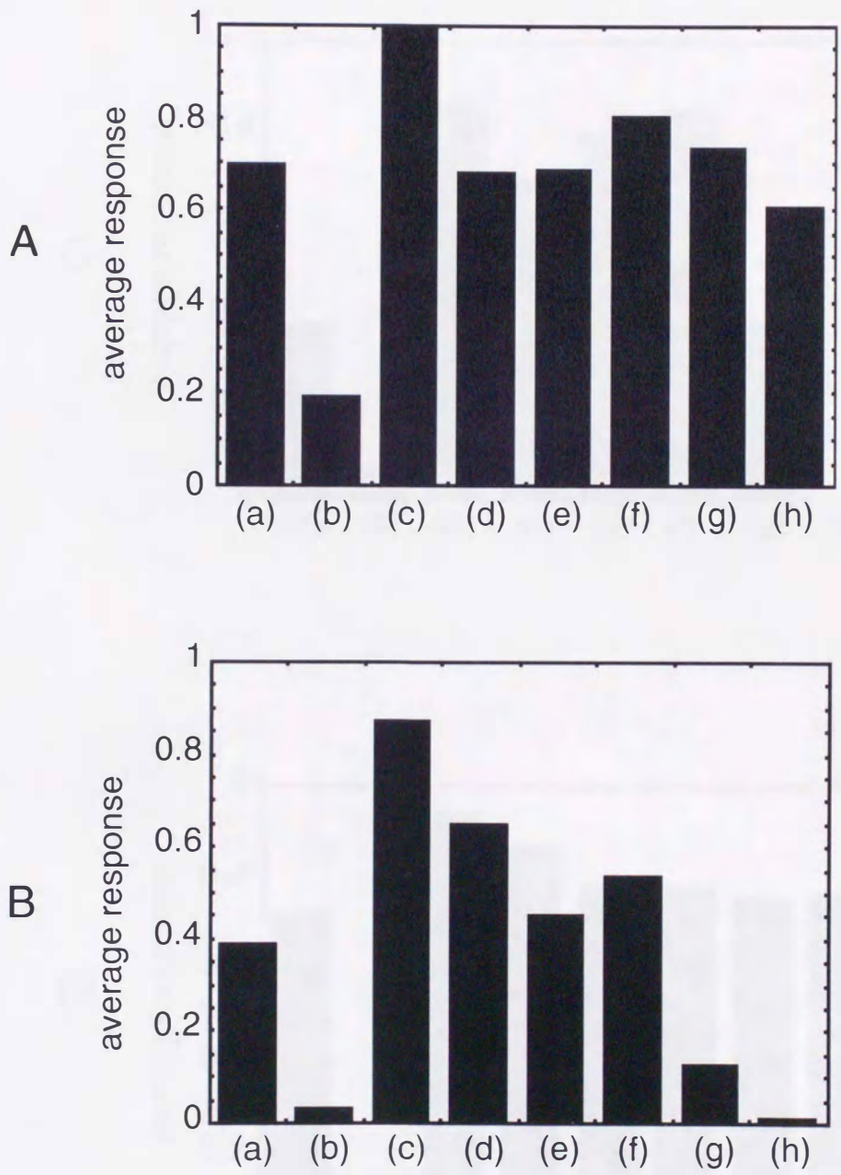
Chapter 1. Figure 4B. M. Kamo



Chapter 1. Figure 5. M. Kamo

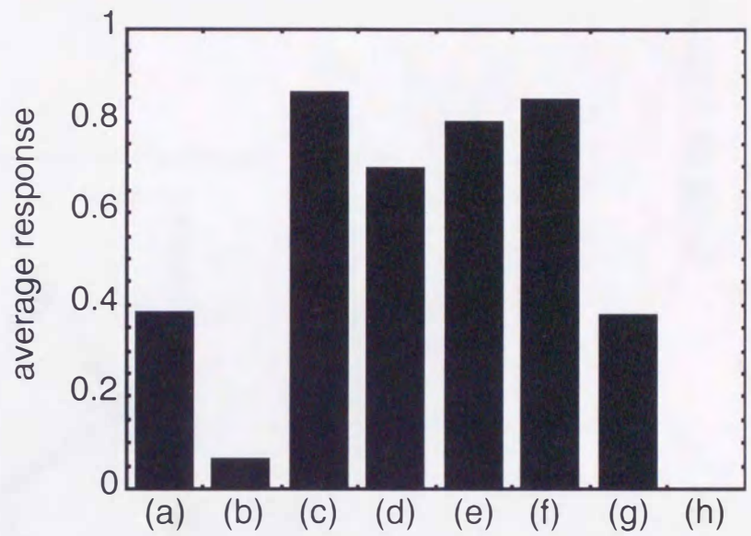


Chapter 1. Figure 6 A&B. M.Kamo

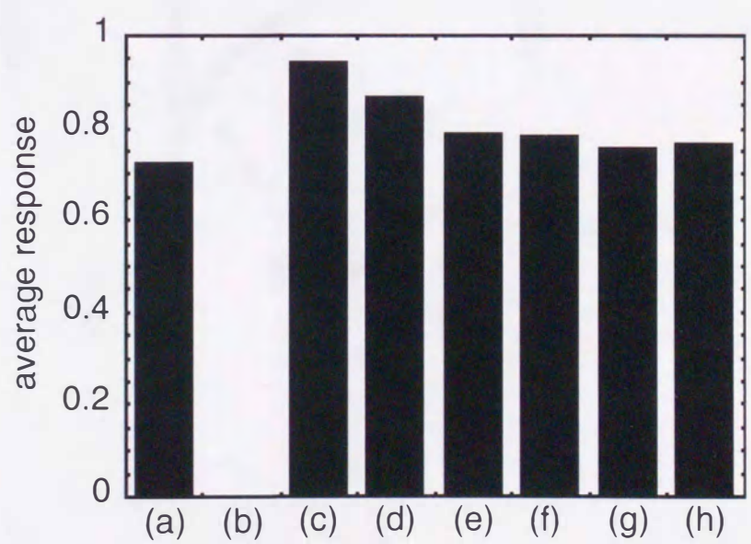


Chapter 1. Figure 6 C&D. M.Kamo

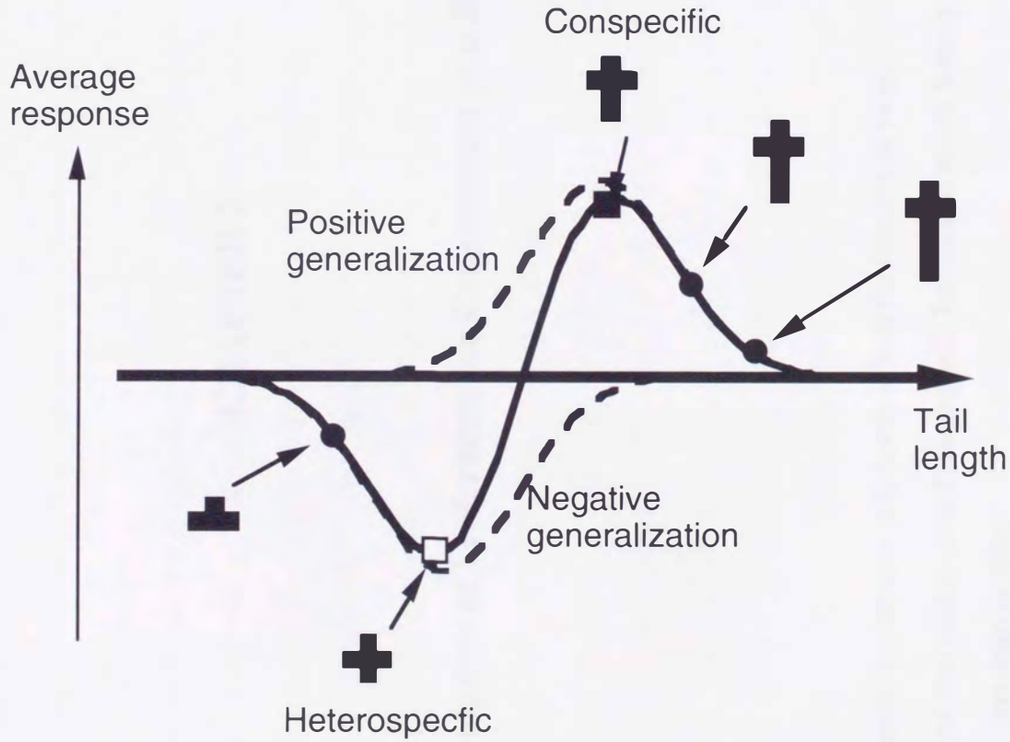
C



D



Chapter 1. Figure 7. M. Kamo



CHAPTER 2

Evolution of Preference for Consonances as a By-product

This chapter was done in collaboration with Professor Yoh Iwasa.

The paper was published in Philosophical Transaction Royal Society of London Series B, (353: 399-460) in 1998.

INTRODUCTION

The female preference for male exaggerated ornaments may have evolved by direct positive fitness effects, or by indirect genetic effects, either through sexy son processes or through handicap principles (see papers cited in Andersson and Iwasa, 1996). Alternatively, the female preference may have evolved as a by-product of evolutionary processes independent of the male trait itself, and male traits may have simply exploited the existing sensory bias of the females (sensory exploitation hypothesis; Kirkpatrick and Ryan, 1991).

When an animal is trained to respond a certain stimulus, it may come to respond to other stimuli that are similar but not exactly the same as the original training signal. This phenomenon is called 'generalization', and is important in understanding the results of learning experiments in animal psychology (Gutman and Kalish, 1956). Generalization is also important in understanding learning by neural network models.

Using a simple neural network model of a female recognition system, Enquist and Arak (1993, 1994, 1998) demonstrated that the preference for a certain set of visual signals can evolve as a by-product of training. The network trained to distinguish some input patterns from others often developed a propensity to favor certain patterns that had not been used in the training procedures (Enquist and Arak, 1993). This result was considered to explain supernormal stimuli in animal behavior and subsequent exaggerated forms of sexually selected male ornaments, or of flower petals (Arak and Enquist, 1993). Kamo *et al.* (1998) studied a slight modification of the training procedures using the same network model as in Enquist and Arak, but they observed a very different outcome. After training, the network evolved to show no sign of supernormal stimuli -- the pattern used for training achieved the highest probability of being accepted by the trained network in all the cases examined. This example urges us to choose carefully the training procedures and network structures when we adopt neural network modeling in studying sensory systems.

A similar model was developed to discuss an inherent preference for symmetric input patterns or simple color (Enquist and Arak, 1994; Johnstone, 1994). Enquist and Arak (1994) argued that colorful or symmetric patterns, which are widely seen in animals or flowers, have evolved because of the sensory bias arising through the coevolution between neural recognition systems and biological signals. Johnstone (1994) also showed that the preference for symmetry occurs as a by-product of selection in the context of mate recognition. However, all of these studies focus on the biased preference for visual signals.

In the present paper, we study the bias in preference for auditory signals. In the theory of harmonics, some combinations of tones are called consonances whereas others are called dissonances (Table 1). Consonances are considered as intervals more comfortable to listen than dissonances. A typical consonance is an "octave", exemplified by a pair of tones with the frequency ratio exactly two. Another example is "perfect fifth" (say "C" and "G") with the frequency ratio close to three halves. Many other examples of consonances have frequency ratios close to the ratios of two small integers (Table 1). In contrast, dissonances tend to have frequency ratios not close to a simple ratio. Why do we feel consonances more comfortable than dissonances? We investigate the question by analyzing simple neural network models and by examining how the preference for certain pairs of tones over others appears in them.

Auditory signals in natural environments are unlikely to be pure tones -- a basic tone (called a fundamental or a root) is almost always accompanied by its harmonic tones. For example, a sound of 220 Hz includes a component of 440 Hz, 660 Hz, 880 Hz etc. In the development of infants, important input auditory signals (e.g. the mother's voice) may be given in different pitches, but in each case the fundamental is accompanied by its harmonic tones. The infant might develop a preference for certain combinations of tones that are commonly included in the harmonic tones of its mother's voice. This may explain the preferences for consonances that have simple frequency ratios (Table 1) as these are more likely to be included in the harmonic tones of input

signals. If so, the preference for consonances and against dissonances can be a by-product of a simpler task, such as to discriminate important sounds (such as mother's voice) from noises. We call this "by-product hypothesis."

Second, the training for a particular consonance and against a particular dissonance may be generalized to a preference for all consonances and against all the dissonances. As explained later, consonances are characterized as sharing low harmonic components unlike dissonances, and hence they may be more similar to each other than dissonances. We call this the "harmonic-overlapping hypothesis."

Alternatively, the preference for consonances over dissonances may be completely arbitrary, and their distinction may be determined by convention or by the historical accident that they were chosen by great composers in old days. If there is no basic reason to favor consonances, we must learn preference for consonances over dissonances, one-by-one. We call this the "taught one-by-one" hypothesis. In the following, we examine the by-product hypothesis and the harmonic-overlapping hypothesis by training neural network models. We treat the taught-one-by-one hypothesis as the null hypothesis so that we favor it if both of the others are rejected.

NEURAL NETWORK MODEL

Auditory signals, or sounds, first arrive at the ear drum, go through the middle ear, and finally reach the cochlea. In the cochlea, sounds are transformed to electric signals by a group of hair cells on the basilar membrane. The hair cells are arranged in the order of their sensitive frequency. Cells on the apex of the basilar membrane react to low frequency sounds, and cells on the base react to high frequency sounds. The system is called "cochlear tuning" (Nicholls, 1992).

In this paper, we model the auditory system as a neural network that has three layers, as illustrated in Fig. 1. An input layer has 37 cells, and cells in the left react to the low frequencies and those on the right react to the high frequencies. We set the ratio of the sensitive frequencies of adjacent input cells at $2^{1/12} = 1.059463$, which corresponds to one semitone in "equal temperament." An octave (the difference of double frequency) includes 12 semitones, and the 37 input cells of the network can cover three full octaves. In the second layer, or hidden layer, there are 10 cells (or neurons), each of which is connected with all the 37 cells in the input layer. Each connection has a weight, which may be modified by training. Finally all the ten cells in the hidden layer are connected to the output layer with only one cell.

The initial values of the connection weights were generated by random numbers uniformly distributed between -1 and 1. The activity level of the jth cell in the hidden layer is:

$$x_j = f(u_j), \qquad (j = 1,...,10), \tag{1a}$$

where

$$u_j = \sum_{i=1}^{37} w_{ji} x_i^{input}. \tag{1b}$$

$f(u_j)$ is a sigmoidal function of u, and is given by $f(u) = 1/(1+\exp[-u])$. w_{ji} is the weight on the connection between the ith cell in the input layer and the jth cell in the

hidden layer. It is positive if a higher activity of the *i*th input cell enhances the activity of the *j*th cell of the hidden layer; but it is negative if the interaction is inhibitory. The output of the system is a sigmoidal function of the sum of signals coming from the hidden layer:

$$x_{output} = f\left(u_{output}\right), \tag{2a}$$

where

$$u_{output} = \sum_{j=1}^{10} w_{oj} x_j . \tag{2b}$$

w_{oj} is the weight for the connection from the *j*th cell in the hidden layer and the output cell.

Starting from the initial network with randomly generated weights, we modified the weights of connections in the direction to make the network show correct responses more often. We trained networks by back propagation, which is a version of the gradient method for optimization (Bose and Liang, 1996). It provides a way to adjust the weights to minimize the sum of the square error between the output of the network and the desirable response (teach signal). A high outputs of a network indicate that the input signal was accepted and low outputs indicate the inputs were rejected.

There is some difference between exact harmonics and the one expressed in equal temperament. For example, perfect 5th is the interval of two tones with frequency ratios of 3/2 which is included in the harmonics of a single tone as the second and the third harmonics. However this is expressed in equal temperament as 7 semitones (see Fig. 2) which is $2^{7/12} = 1.498$, which is close to but not exactly the same as 3/2. In this paper, we assume that generalization in the space of frequency works, and training with tones of frequency ratio 3/2 enhance the reaction of the auditory system to the intervals of ratio 1.498.

BY-PRODUCT OF TRAINING

Input signals in natural environments are composed of a basic tone ('fundamental') and its harmonic tones. Figure 2 illustrates all the differences between harmonic components of a monotone shorter or equal to an octave. The numerals indicate the number of semitones included in each interval. All these intervals correspond to consonances but most intervals that do not appear in Fig. 2 are dissonances (see Table 1). The only exception was minor 6th (8 semitones) that is classified as a consonance (Shimofusa, 1972; see Table 1), but not included in Fig. 2. Since several low harmonics are usually more important than harmonics of still higher frequencies, we assumed that each basic tone is accompanied by six lowest harmonic tones with equal amplitude.

The networks were trained to accept monotones accompanied by harmonic tones shown in Fig. 2 and to reject noises. Each accept signal was shifted for all the cases in which two or more harmonic tones were included in the audible range of 37 input cells. There were 52 input signals in total.

As a simple model for random noises to reject, we generated uniform random numbers between 0 and 1 to all the cells in the input layer. In each step of weight modification, the network was presented 50 accept signals and 500 reject signals.

The training continued until the average error rate became less than 5%. After the training period, intervals of two notes within an octave (a difference less than or equal to 12 semitones) without harmonic tones were given to the network. There were 366 test patterns, and 211 of them were consonances, whilst the other 155 patterns were dissonances. If networks have no systematic preference for a group of inputs over the others, we expect that the average response of independently trained networks to the former is higher than the average response to the latter with 50 % chance. Hence we can apply a binomial test to detect a systematic bias.

Figure 3 illustrates the average responses of the trained networks to the consonances and to the dissonances defined by music theory (table 1). The former was

larger than the latter for 46 networks among 50 replicates, which was statistically significant (binomial test, $z = -5.9397$, $p < 0.001$).

Null intervals (the relationship with the same tone) were excluded from the analysis in Fig. 3, although they are consonances (perfect 1st). If these were included, we had a much larger difference in responses to the two groups of input intervals than in Fig. 3.

This result supports the by-product hypothesis.

GENERALIZATION BETWEEN CONSONANCES BY HARMONIC OVERLAPPING

The second issue is whether the preference for a particular consonance is generalized to other consonances. If two fundamentals are consonances, the frequency ratio is a ratio of small integers, and hence their sets of harmonics tend to share common tones. Sharing the common harmonic tones is less likely to occur if two fundamentals are dissonances. Using the propaty of having common harmonic tones, the network may be able to distinguish the two classes of intervals. If one consonant pair of tones is favored in training, the network might develop the preference for other consonances that did not appear in training.

To test the possibility of generalization, we used accept signals of intervals of 7 semitones (perfect fifth), and reject signals of 6 semitones (diminished fifth) (Table 1). We used all the cases in which the fundamentals of both input signals were included in the first octave. Each network experienced 13 input patterns in each trial (6 pairs of tones with perfect fifth and 7 pairs of tones with diminished fifth in an octave).

Figure 4 illustrates the average response to consonances and the average response to dissonances that were not used in the training. The networks were trained with the tones with harmonics, but were tested by the tones without harmonics. The average response to consonances was not higher than the average response to dissonances (55 times among 100 replicates, statistically not significant, binomial test, $z = 1.0$, $p = 0.097$). This implies that the trained preference was not generalized to other consonances.

DISCUSSION

We examined how the preference for some pairs of tones might arise. A special feature of the auditory system, compared to visual, olfactory, or other sensory systems, is that input signals are almost always accompanied by their harmonics. Characteristic biases commonly occurring in auditory signals might be explained as a coincidental outcome of this feature.

Generalization is an important property of all neural systems, including auditory systems. This is also important in forming preference for consonances. For example, the input signals includes harmonics 3 times higher in frequency than the basic tone, and is exactly 1.5 times of the first octave, thus causing a preference for the perfect fifth. However the perfect fifth, expressed in terms of equal temperament or 7 semitones, is $2^{7/12} = 1.4983$, which is close to but not exactly the same as 1.5. The latter is just intonation. However, thanks to the generalization ability of the auditory system, the network can develop a preference for 7 semitones from the input signal. The difference between just intonation and the closest interval of equal temperament may be larger than this case, but is still small (Table 1).

In our trials, networks trained to accept monotones with harmonics and reject random noises developed a preference for consonances spontaneously. Hence, preference for consonances can be acquired without being taught one-by-one. It is simply a by-product of training procedures to distinguish monotones from random noises, caused by the fact that most auditory signals are accompanied by harmonic tones with frequency equal to multiples of the basic frequency. We would like to conclude tentatively that the preference for consonances arises as a by-product of training for monotones against random noises.

On the other hand, we found that the networks trained to accept perfect fifth and to reject diminished fifth did not develop a preference for consonances that were not used in the training (Fig. 4). This result implies that the training to favor a particular

consonance and to reject a particular dissonance was not generalized to other consonances or dissonances.

In summary, the preference for consonances is likely to have arisen inadvertently. This suggests that preferences between auditory signals may not be completely conventional or arbitrary. If so, the same process should work for animals other than human that have a similar auditory range, and they too might find consonances more comfortable to listen to than dissonances.

REFERENCES

Andersson, M. and Y. Iwasa. 1996. "Sexual selection". *Trends Ecol. Evol,* **11**: 53-58.

Arak, A. and Enquist, M. 1993. Hidden preferences and the evolution of signals. *Phil. Trans. R. Soc. Lond. B*, **340**: 207-213.

Bose, N.K. Bose and Liang, P. 1996. *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. New York: McGraw-Hill Inc.

Burns, E. M. and Ward, W. D. 1982. Chapt. 8. In *The psychology of music.* (D. Deutch ed.) (Japanese Translation), pp. 301-334. Tokyo: Nishimura Shoten Co.

Enquist, M. and Arak, A. 1993. Selection of exaggerated male traits by female aesthetic senses. *Nature*, **361**: 446-448.

Enquist, M and Arak, A. 1994. Symmetry, beauty and evolution. *Nature,* **372**: 169-172.

Enquist, M. and Arak, A. 1998. Neural Representation and the Evolution of Signal Form. In *Cognitive Ecology* (R. Dukas ed.), pp. 21-87. Chicago: The University of Chicago Press.

Gutman, N. and Kalish, H. I. 1956. Discriminability and stimulus generalization. *J. Exp. Psychol.*, **51**: 79-88.

Johnstone, R.A. 1994. Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature,* **372**: 172-175

Kamo, M., Kubo, T. and Iwasa, Y. 1998. Neural network for female mate preference, trained by a genetic algorithm. *Phil. Trans. R. Soc. Lond. B*, **353**: 399-406.

Kirkpatrick, M. and Ryan, M.J. 1991. The evolution of mating preference and the paradox of the lek. *Nature* , **350**: 33-38.

Nicholls, J.G. Martin, A. R. and Wallace, B.G. 1992. *From neuron to brain* 3rd ed. Massachusetts: Sinauer Associates, Inc. Publishers.

Shimofusa, K. 1972. *Harmonics. (Waseigaku).* pp. 157 (in Japanese). Tokyo: Ongakunotomosha Publ. Co.

Table 1.

Harmonistic definition of intervals. Frequency ratios are based on equal temperament, in which the interval with n semitones has frequency ratio of $2^{n/12}$. Those intervals with this frequency ratio close to a ratio of two simple integers are consonances and others are dissonances. Ideal ratios are based on "just intonation" (Burns and Ward, 1982), and classification of consonance/dissonance is based on Shimofusa (1972).

		interval name	number of semitones	frequency ratio	ideal ratio
consonances	perfect	perfect 1st	0	1.0	1:1
		perfect 8th	12	2.0	2:1
		perfect 5th	7	1.498	3:2
		perfect 4th	5	1.335	4:3
	imperfect	major 3rd	4	1.260	5:4
		minor 3rd	3	1.189	6:5
		major 6th	9	1.682	5:3
		minor 6th	8	1.587	8:5
dissonances		major 2nd	2	1.122	9:8
		minor 2nd	1	1.059	16:15
		augmented 4th	6	1.414	45:32
		diminished 5th	6	1.414	64:45
		major 7th	11	1.888	15:8
		minor 7th	10	1.782	16:9

Figure Legends

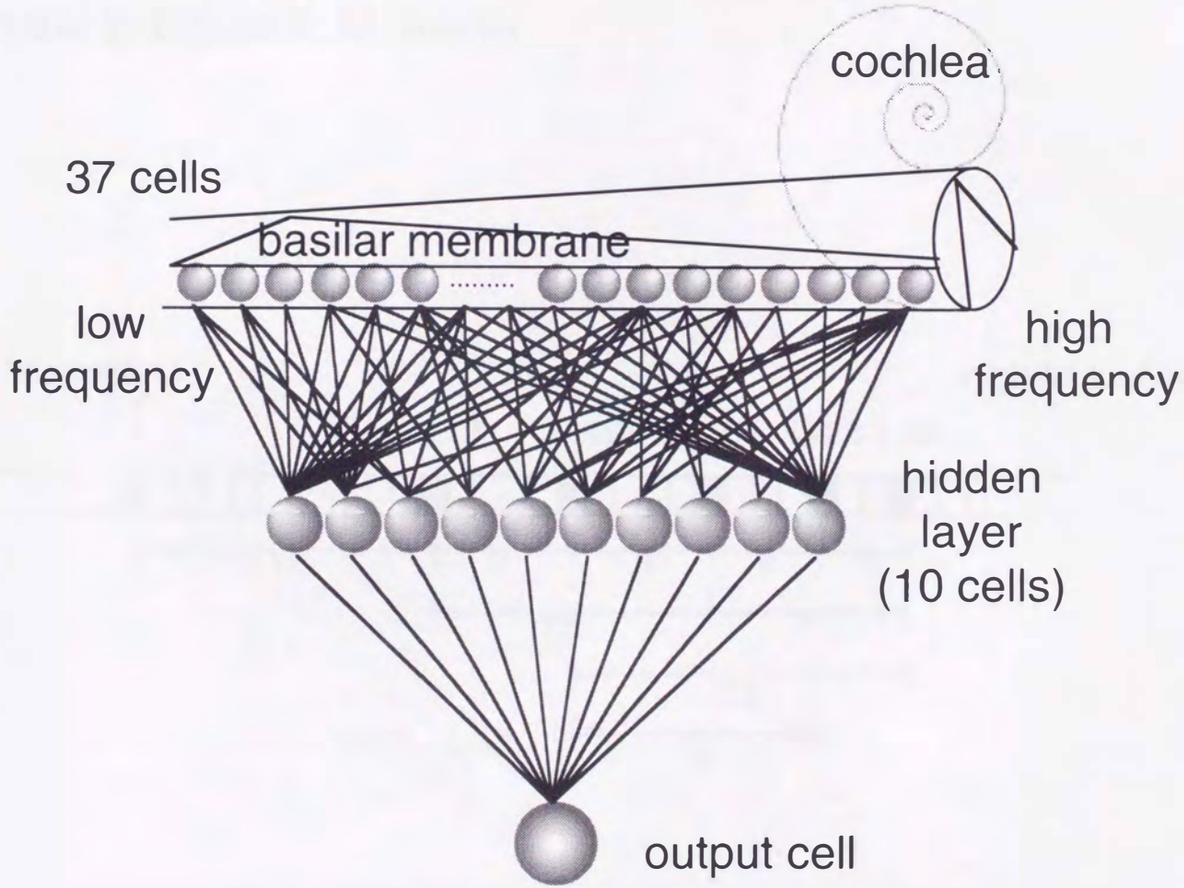
Figure 1 The auditory system and the neural network used in this study. Each cell has a sensitive frequency, and cells are arranged in the input layer according to the frequency (cochlea tuning). The left side of the input layer respond to the lower frequencies and right side respond to the higher frequencies. Music intervals between adjacent cells are a semitone, and the musical range of this network is three octaves.

Figure 2 Intervals between harmonics accompanying a single basic tone. *i* indicates the fundamental, and the other black squares are for the harmonic tones. The numerals indicate the length of intervals, given as the number of semitones included. These intervals are all consonances.

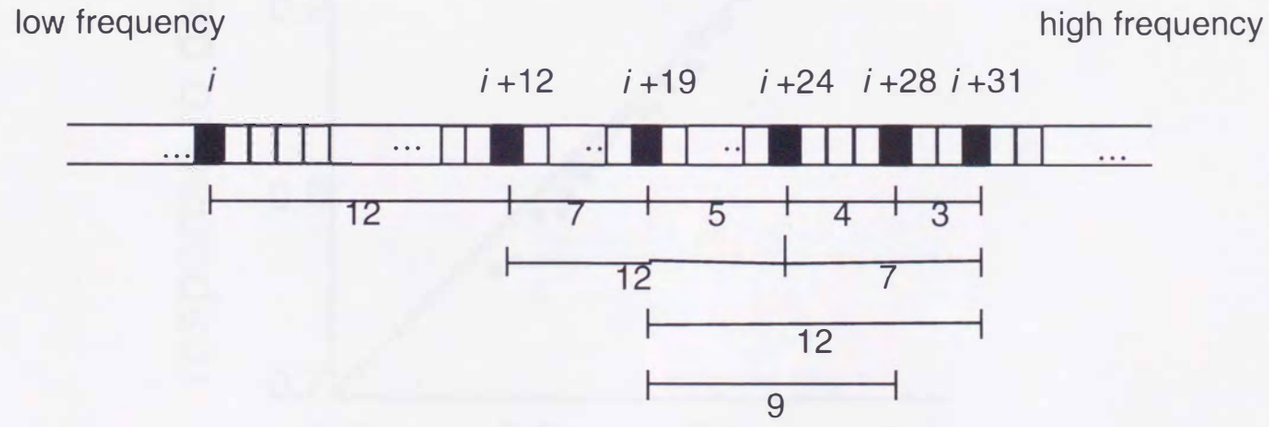
Figure 3 The average responses of 50 trained networks trained independently. The average responses to consonances and those to dissonances, defined in Table 1. The average response to consonances was higher than that to dissonances (binomial test, $z = -5.9397$, $p < 0.001$). The line has slope 1, intercept 0 and is not the regression.

Figure 4 The average responses of 100 independently trained networks in harmonic-overlapping hypothesis. The networks were trained to favor perfect fifth (7 semitone intervals) and to reject diminished fifth (6 semitones). Test signals were intervals without harmonics. Out of 100 networks, only 55 networks showed the preference to consonances not used in the training, which is not statistically significant (binomial test, $z = 1.0$, $p = 0.097$). The line as in Figure 3.

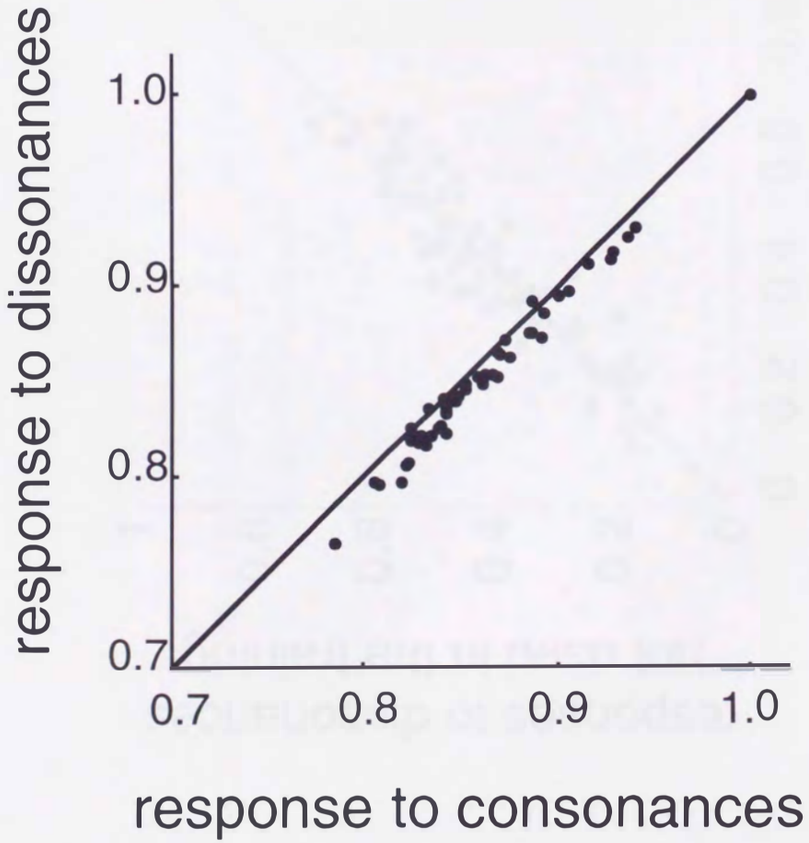
Chapter 2. Figure 1. M. Kamo



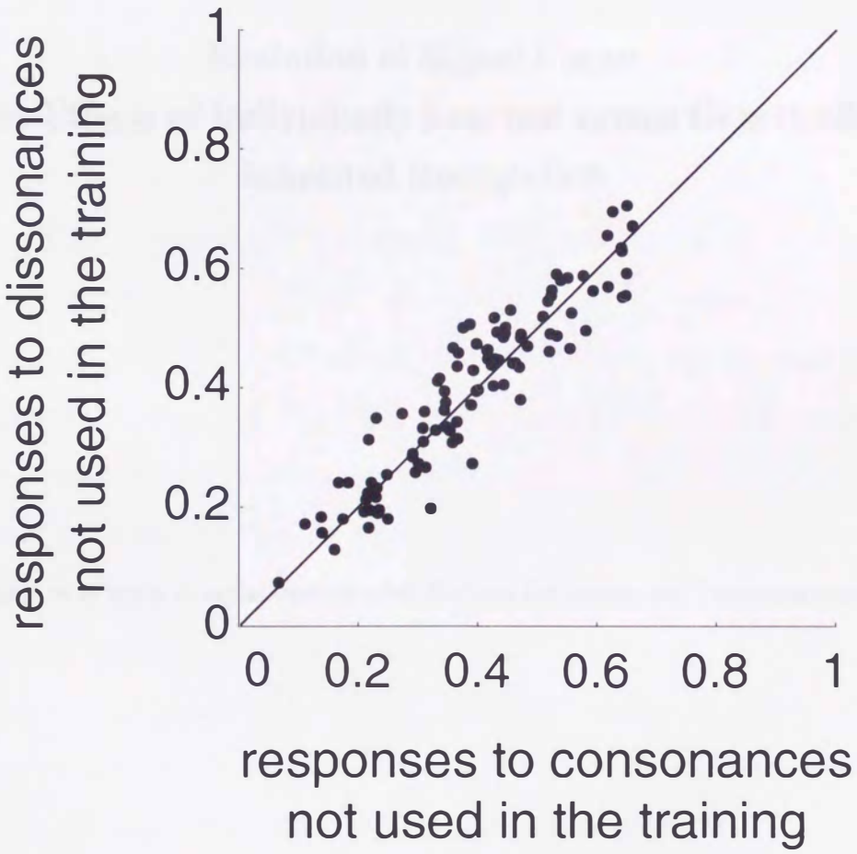
Chapter 2. Figure 2. M. Kamo



Chapter 2. Figure 3. M. Kamo



Chapter 2. Figure 4. M. Kamo



CHAPTER 3

Evolution of Signal Form: The Effects of Individually Learned versus Genetically Inherited Recognition

This chapter was done in collaboration with Stefano Ghirlanda and Professor Magnus Enquist.

INTRODUCTION

The evolution of signal form has recently received a lot of interest (see reviews by Enquist & Arak 1998 and Bradbury & Vehrencamp 1998). Attention has been given to determinants of form including coding and transmission of information, receiver bias and strategic factors. Here we investigate another determinant potentially important for signal form. Recognition of a signal may be genetically inherited or established by some learning process. Here we focus on individual learning in which each individual must learn based its own experiences. In genetic recognition the receiver inherits experiences obtained during the species evolutionary history. These differences may have effect on the evolution of signal form because mistakes occur in the former cases but not in the latter one.

Both genetic and learned recognition occurs in nature. Frog females recognize the call of their conspecific males without any learning. Similarly snakes recognize food without learning (Arnold 1981). Examples in which learning is fundamental include mate and sex recognition zebra finches and predators learning to avoid unsuitable prey. Why this diversity exists is an important issue but not the topic of this paper. Rather we ask whether individually learned and genetically inherited recognition have different consequences for the evolution of signal form. In particular, we are interested in the degree of exaggeration and cost of the signal.

Most studies of signaling seem to assume genetically inherited recognition. The only area in which the effect learning has been extensively discussed is the evolution of aposematic coloration (see e.g. Guilford 1990, Leimar et al. 1986).

To proceed, we need to understand the consequences of mistakes for the players involved. We assume that signal senders and signal receivers both benefits from the receiver accurately recognizing the senders. Finesses depend from a number of encounters between senders and receivers. A sender's fitness depends on the behavior of the receiver, not on what the receiver learns from the encounter (unless the same sender and receiver meet again or senders are related), since learning will affect only

future behavior. Thus, increased rate of learning can not be the driving force, instead we have to seek the driving force in the receivers response (Leimar et al. 1986; Guilford 1990).

Previous studies of signal form have pointed the importance of conflict (Grafen 1990; Arak & Enquist 1995). Typically this conflict is present in the pay off matrix, for instance, the sender benefit from stronger response than the receiver. However, conflict can also be due to lack of knowledge. The distasteful bug and the predator benefit both from the predator not attacking. But as long as the predator cannot discriminate between edible and non-edible bugs there is a conflict that can only resolved by the predator trying different bugs. Once the predator can discriminate the conflict is gone. It follows that in the case of genetically inherited recognition the conflict disappears once discrimination is established. In the case individually learned recognition the conflict reappear in each naive receiver.

In this paper we study theoretically evolution of signal form under both genetically inherited and individually learned recognition. We simulate coevolution between senders and receivers using artificial neural networks as the receivers recognition mechanism.

MODEL

We used a three-layered neural network with 10 receptors in an input layer, 5 cells in a hidden and an output cell. The cells between layers are fully connected with a certain weights. Initially these weights are randomized by the uniformal random numbers between −1 to 1. Signals used in this study have 10 stimulus elements and the same number as the network's receptors. The networks are trained to have correct outputs for two learning tasks. One is an ignored signal which is called s^-. The stimulating level of all elements in s^- is 0. Since the signal does not stimulate anything, we can comprehend that the color of the signal is black. The other is the coevolving signal (s^+). The initial stimulating level of all elements in the signal is 0.02 and the signal evolves by mutating the stimulating level of each element. The targeting point to s^- is 0.09 and that to s^+ is 0.9. A network is trained to have those responses to the signals.

The fitness of signal and the network are:

$$F_N=(1-|T_{s^+}-O_{s^+}|)\times(1-|T_{s^-}-O_{s^-}|) \qquad (1)$$

$$F_S=(1-|T_{s^+}-O_{s^-}|) \qquad (2a)$$

, or when considering the costly signal:

$$F_S=(1-|T_{s^+}-O_{s^-}|)-Cost \qquad (2b)$$

where F_N and F_S are a fitness of a networks and signal respectively. T_i is a targeting point of object i ($i=s^+, s^-$), and O_i is an output (or response) to the learning task i. We defined that the signal costs are squared sum of each unit's stimulating level. That is

$$Cost=\sum_i x_i^2/n \qquad (3)$$

where x_i is a variable of i-th unit's stimulating level in the signal and n is a number of units in the signal. It is ten in this study. We defined that the signal which has high stimulating level is exaggerated.

Two evolutional mechanisms are simulated. One is the genetically inherited recognition. Since behavior of individuals is designed by genes, children have the same behavior as their parents. The other is learned recognition; their behavior is influenced by the

individuals' experiences. Those two differences are modeled as follows.

In genetic recognition model, there are two networks and two s^-. One is a original which is inherited from the last generation, the other one is an its mutant. All signals are presented to all networks, and the better signal and the better network are inherited to the next generation. This procedure is clearly mentioned in Enquist & Arak (1993, 1994) or Kamo et al. (1998). In learning recognition, in every generation, networks connection weights are initialized. Then the network learns to achieve correct responses to each learning task for a certain learning period. Even when the network responses to each signal do not reached to its targeting, network's learning is stopped after a certain training period. After the training period is over, mutant signal is created and both original signal and the mutant signal are shown to the network. The fitnesses of the signals are calculated using the Eq. 2, the better signals are transmitted to the next generation. In the next generation the network is initialized again. In this procedure, although the signal is still in a genetic transmission, the behavior of the network differs generation by generation because they have a different experience. For network's individual learning, we adopted the back propagation. The back propagation is a standard learning algorithm for multi layer (Hykin 1999). The weights modification is done gradually to minimize the differences between teach signal (response target) and the present output of the network. When the network's learning period is longer, the network's response becomes more precise. We can expect that the network's responses after learning mainly depend on two factors. First one is the length of the learning period. If network can keep learning to that signal for long time, the response to the signal would be precise. Second one is the signal form. If the signal had an easy form to learn, networks can obtain precise output to the signal in short time. To examine the idea, before we do evolutional simulation, the learning speed of the network is investigated.

Learning curves

Figure 1 shows the networks learning speed of the network in various learning tasks.

For all trajectories, s^- is fixed as the black signal (all stimulating units equal 0). The solid line shows the learning trajectory when all elements in the s^+ signal are consisted by stimulating level of 1.0 (white signal). In this case, the responses to each signal converge to their targets quickly. On the other hand, all elements in the s^+ are 0.1 the speed to converge to the target is very slow. We can conclude that when s^+ and s^- is similar, the learning speed is slow and when these two signals are different, the network's learning speed is fast. For all cases, when we obtain a longer learning period, the network's discrimination ability becomes more precise. From Fig. 1, we can expect that the signal will evolve to the different form from s^-. Our another interest is that how the learning period affects the evolution of the signal form. When the networks is allowed to learn only for 100 back propagation cycles, what form will the signal evolve to? When the networks is allowed to learn 200 back propagation cycles, what form will the signal evolve to?

RESULT

Cost Free Signal.

Evolutionary trajectories of four cases are shows in Figure 2. Generally, the simulations reached a steady state after a number if generations. In all figures, the vertical axes indicates fitness of signal (solid line), the exaggeration in the signal (solid line with a black circle), networks output to s^+ (dashed line) and the network's output to the s^- (solid line with "x"). In the case of genetic recognition, 10 to 30 thousands generations were needed to reach to the optimal responding. The signal evolved to highly exaggerated form. In the case of learned recognition in, the steady state was reached within less than 1,00 generations. The number of back propagation cycles in a generation is 500 (see figure 1). In this case the network's response did not reached to the optimal responding. The signal evolved to the most exaggerated form (all stimulating units in the signal become 1). To obtain the optimal responding, networks have to obtain more back propagation cycles in a generation.

In both cases signals evolved to exaggerated form immediately. The result shows that the signal will evolve to less similar form from the black signal (s^-).

Costly Signal.

The right figures in Figure 2 shows the evolutional trajectories when the exaggeration is costly. Generally the exaggeration in signals were not so high as in the non-costly cases. In inherited recognition, optimal responding is attained within 30 to 40 thousands generations. However the fitness of the signal did not reached to the maximum. This is because that even when the evolution reached in the steady state, the signal remained in the costly form. The bottom figure in the right column shows the case of learning case. The back propagation cycles in a generation is 500. The optimal responding did not attained. The signals were remained slightly costly.

In these four figures, a mean of ten simulations is shows.

Comparisons.

In all cases, when the exaggeration is costly case and not costly case, the exaggerated signal evolved. However the degree of exaggeration is different, in the case of the exaggeration of the signal is costly and not costly. This is apparent that the signals can evolve to the most exaggerated form when exaggeration is not costly. The main topic in this study is to ask if there is a difference between learned recognition and inherited recognition. Figure 3 illustrates the comparison of exaggeration in signals after the simulation reached in the steady state comparing both the learned recognition and the inherited recognition. In all cases, mean of ten simulations are shows. The exaggeration in the case of inherited recognition is shown in a dashed line. In the case of learned recognition, some simulations were done in various back propagation cycles. When more back propagation cycles are obtained, the network's responses are more precise (Fig. 1). When small back propagation cycles are obtained, the exaggeration is the signal is very small. However, the back propagation cycles in a generation comes to a certain number (more than 200 back propagation cycles), the exaggeration in signals become high, even higher than the case of inherited case. More back propagation cycles are obtained, the exaggeration in signals become smaller, but still higher than the case of genetic case.

DISCUSSION

In this paper, we studied the evolution of the exaggeration form in signals comparing the genetically inherited recognition and the individually learned recognition. In both cases, when the exaggeration is not costly, exaggerated signals easily evolved. Why the exaggeration in signal evolved? When the signals had the exaggerated form, networks could learn to have correct responding to those signals quickly (Fig.1). This is why the signals evolved to the exaggerated form. When the exaggeration is costly, the evolution of the signal is depending on two factors, output from the network and the exaggeration costs. For neural networks, they favor the exaggerated form, since those signals are easy to learn. For signals, they also favor the exaggerated form to obtain the more correct responding form the network. However, on the other hand, too much exaggeration is not favored because it is costly. These two conflicting evolutional forces decide the evolution of the signal forms. In all simulations, the evolution reached in a steady state as illustrated in Figure 2. The exaggeration in the signal after simulations reached in the steady state are shown in Fig. 3.In Fig. 3, the exaggeration in signal become smaller when the more back propagation cycles are obtained. Why do they become smaller? When the network do not learn so much, signals will be in more exaggeration to compensate for smaller outputs of the network. The signals should evolved to some easier form for networks to learn. The form is exaggeration.

To examine this idea, we investigate the evolution of the signal in various cost intensities and various back propagation cycles. Now the signal costs are:

$$Cost = I \times \sum_i x_i^2 / n \qquad (4)$$

where I is a cost intensity. Results are shown in Figure 4. Fig. 4A illustrates the exaggeration in a signal. Horizontal axis shows the cost intensity (I) and the vertical axes shows the back propagation cycles in a generation and shown in a rough log scale. White area indicates that the signals are in the exaggerated form. Those white signals are seen where I=0 (cost free). However, when cost intensity is more than 0, signal did not evolved to the most exaggerated form, but exaggeration is somewhere in the

middle. When the networks do not learn so much (less back propagation cycles in a generation), exaggeration in the signal is very small. The networks learn more in a generation, signal evolved to exaggerated form and more learning leads to less exaggerated form. This tendency is seen all field in the Fig. 4A. This is because when the network learn enough, the small difference between s^+ and s^- can be discriminated. Then the signal becomes the less exaggerated form obtaining the same fitness. The fitness of senders is shown in Fig. 4B. The exaggeration becomes less where cost intensity is big and enough learning field in the Fig. 4A, but the fitness of the senders are not become small.

Why the signal in learning recognition is more exaggerated than that of genetic recognition? The driving force for exaggeration is, first the exaggerated signals are easier to learn (Fig. 1). When the network's learning period is restricted, the exaggerated signals would get higher output from the network. Second, the network's response gradients are biased to the exaggerated form because of the individual learning, then a mutant which has more exaggerated form would have higher response from the network. The second idea is illustrated in Figure 5. The network's response after the learning is peak shifted towards to the exaggerated form. The cost becomes higher in more exaggerated form. The evolution of the signals stop when the higher response caused by the peak shift and the signal costs are balanced.

REFERENCES

Arak, A. & Enquist, M. (1995). Conflict, receiver bias and the evolution of signal form. *Phil. Trans. R. Soc. Lond.* B. **349**, 337-344.

Bradbury & Vehrencamp (1998) *Principles of animal communication.* Sinauer Assoc.

Enquist, M. & Arak, A. (1993). Selection of exaggerated male traits by female aesthetic senses. *Nature, Lond.* **361**, 446-448.

Enquist, M. & Arak, A. (1994). Symmetry, beauty and evolution. *Nature, Lond.* **372**, 169-172.

Enquist, M. & Arak, A. (1998). Neural Representation and the Evolution of Signal Form. In: *Cognitive Ecology* (R. Dukas ed.) Chapt. 2. p. 21-87. Chicago: The University of Chicago Press.

Guilford, T. (1990) The secrets of aposematism – unlearned responses to specific colors and patterns. Trends Ecol. Evol. **5**. 323-323

Grafen, A. (1990). Biological signals as handicaps. J. Theor. Biol. **144**, 517-546

Leimar, O., Enquist, M., & Sillen-Tullberg, B. (1986). Evolutionary stability of aposematic coloration and prey unprofitability: A theoretical analysis. *Am. Nat.* **128.** 469-490

Hykin, S. (1999). *Neural Networks.* Ontario: McMaster University.

Kamo, M., Kubo, T. & Iwasa, Y. (1998). Neural network for female mate preference, trained by a genetic algorithm. *Phil. Trans. R. Soc. Lond.* B. **353**, 399-406.

FIGURE LEGENDS

Figure 1. The learning curves of the networks used in this study. In both figures, horizontal axis shows number of back propagation cycles, the number of times for weights modification. In one simulation, there are two learning tasks. One is ignored signal, usually called as s^-. The signal has ten stimulating elements. The stimulating level of the units is 0, corresponding the black signal. The network's responding level to the signal is 0.1, it is trained to have the output when the signal is shows. The other is a accept signal, which is usually called as s^+. Network's responding level to the signal is 0.9. The learning speed in various s^+ is simulated. The upper figure shows the responses to s^+ and the other shows the responses to s^-. In this simulation, four different s^+ are used. The solid line shows the learning curve when the all elements in the signal have stimulating level of 1.0. Dashed line shows 0.5, line with black dots shows 0.2 and the line with "+" shows the 0.1 signal. In all cases, the output to each signal goes to its target when the network learns more. However, the speed to go closer to the target is very slow when the s^+ is 0.1. Contrary, the signals is consisted by stimulating level of 1.0, the speed is fast.

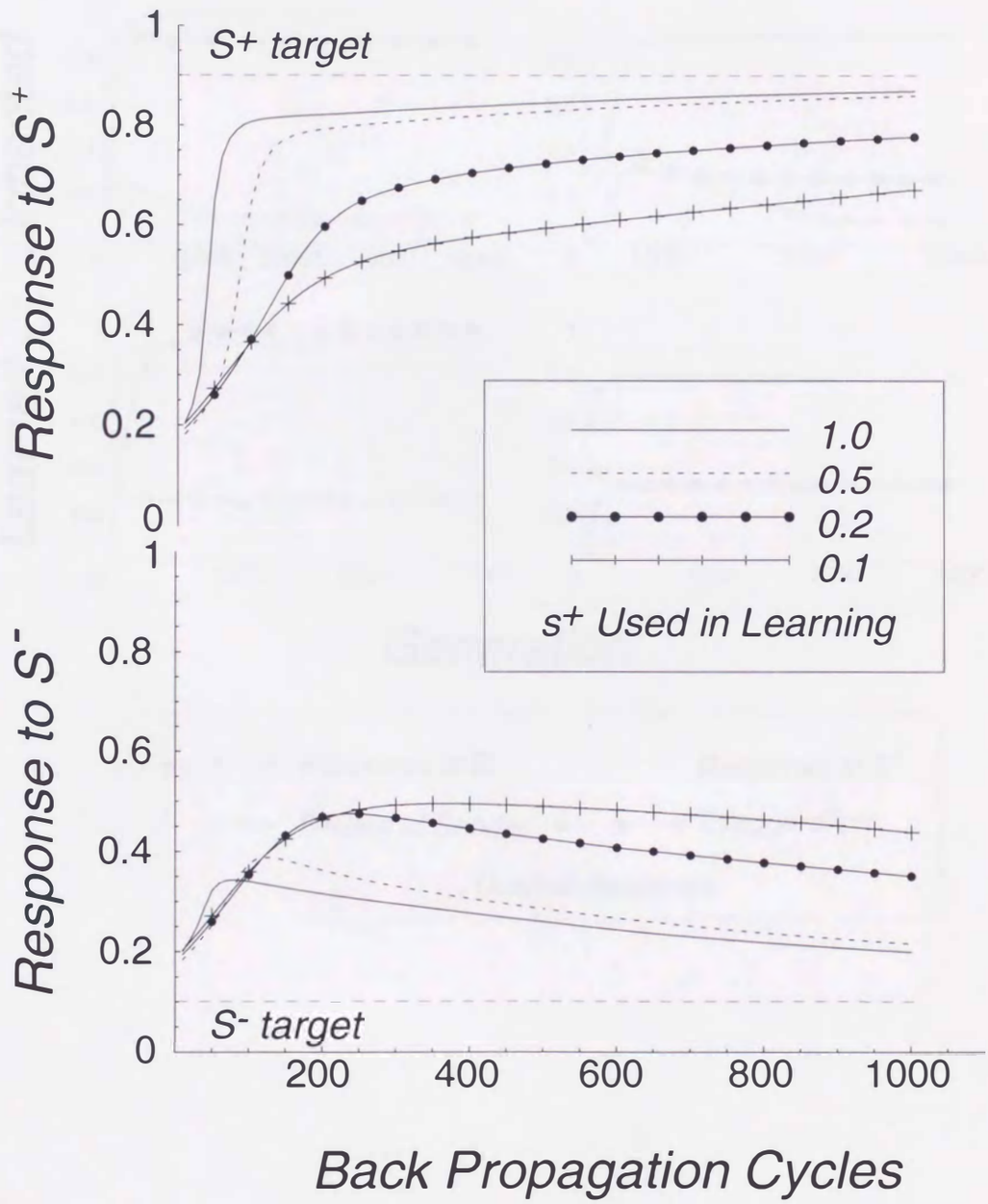
Figure 2. Evolutional trajectories are shown in costly case versus non-costly case and inherited case and learned recognition. The four trajectories are shown in each chart. Solid line shows the fitness of the signal, the line with "x" shows response to s^-, dashed line shows the response to s^+ and the line with black circle shows the exaggeration in the signal. In general, all the simulation is in a steady state after number of generation. To reach to the steady state, 20 to 40 thousands generations are need in the case of inherited recognition. In the case of learned recognition, only hundreds generations are need. The evolution of exaggeration is very high when the exaggeration is not costly. Even in the costly case, signals are slightly exaggerated. In all figures, the mean of ten simulations is shown.

Figure 3. The signal costs after the evolution of signal reached in a steady state. Horizontal axis shows the back propagation cycles in a generation and the vertical axis shows the exaggeration in the signal. A mean of ten simulations is shown. Dots show the mean of exaggeration, and bars show the standard errors. The exaggeration in signal in genetic cases is shown in solid line with two dashed lines for comparison. When networks learning in a generation is very small (less than 100 back propagation cycles), exaggeration in signal is very small, even smaller than that of genetic case. Exaggeration in signal is the highest around 200-300 back propagation cycles. When networks learning is more precise, the exaggeration became lower.

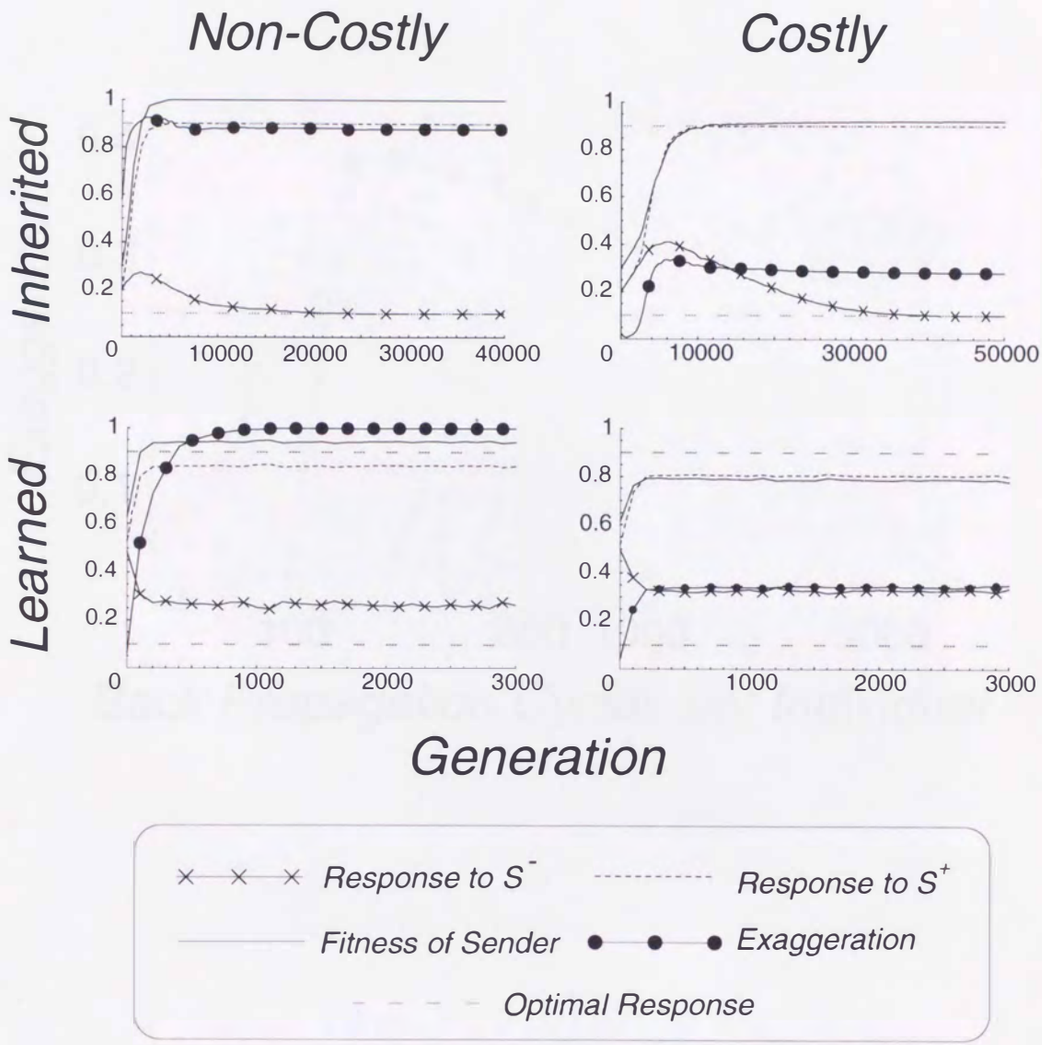
Figure 4. A shows the exaggeration in signal. Horizontal axis shows the cost intensity (I). Vertical axes shows the back propagation cycles in one generation and is shown in quasi log scale. White indicates that the signal is not exaggerated and the darker area shows that the signal is exaggerated. When the cost intensity is zero, the signal is mostly exaggerated. The tendency when the network's learning is more precise, the signal is less exaggerated can be seen in all area. B shows that the fitness of the signal in various cost intensity and back propagation cycles. In the area in right above, the exaggeration in signal was smaller than the other area but the fitness of the sender is not small.

Figure 5 The mimetic diagram of the evolution of exaggeration. The horizontal axis shows the exaggeration in signal and the vertical axes shows the response of the network. The solid curve shows the network's response curve after learning and the dashed line shows the signal cost. The response curve after learning is expected to be biased towered the exaggerated form. When the fitness of original signal and the mutant are compared, if the mutation occurred towered the exaggerated form, the signal has an advantage because it can get higher response from the network. However, too much exaggeration leads to high cost. Then the evolution stopped where these two evolutional forces are balanced.

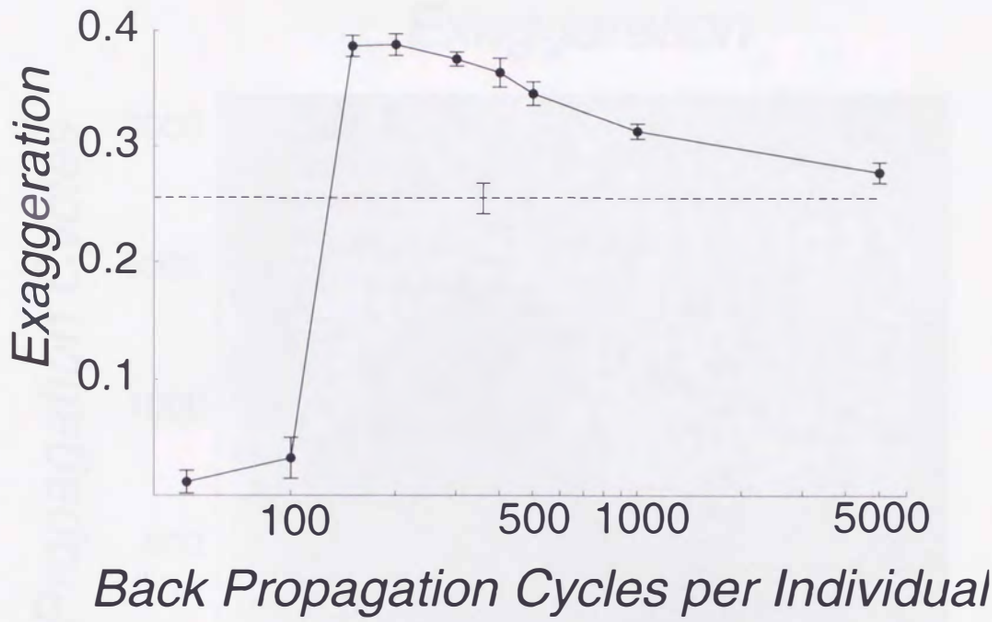
Chapter 3. Figure 1. M. Kamo



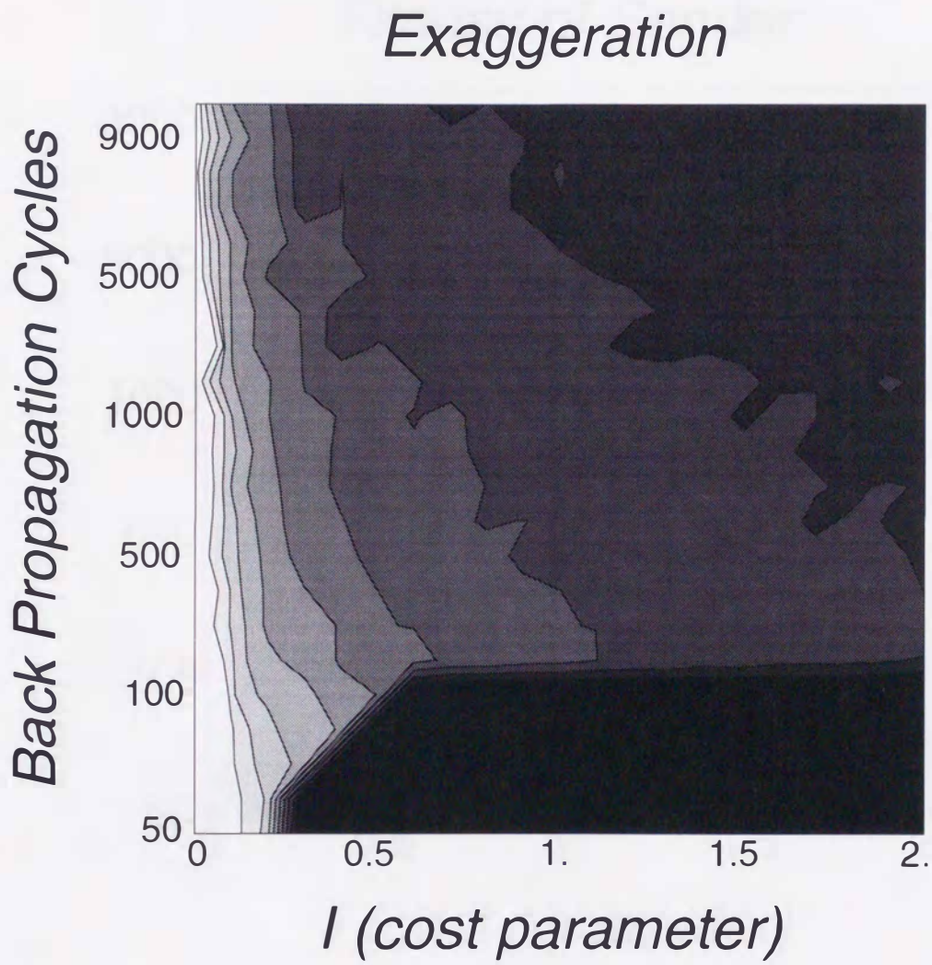
Chapter 3. Figure 2. M. Kamo



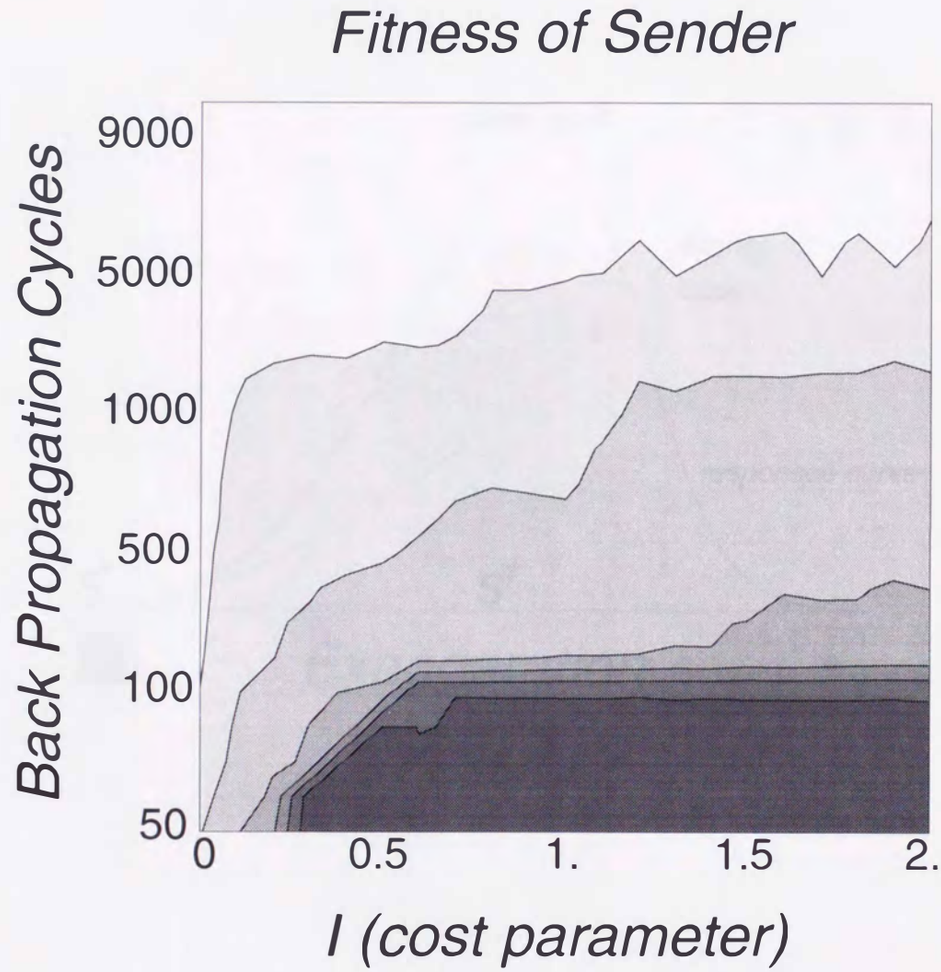
Chapter 3. Figure 3. M. Kamo



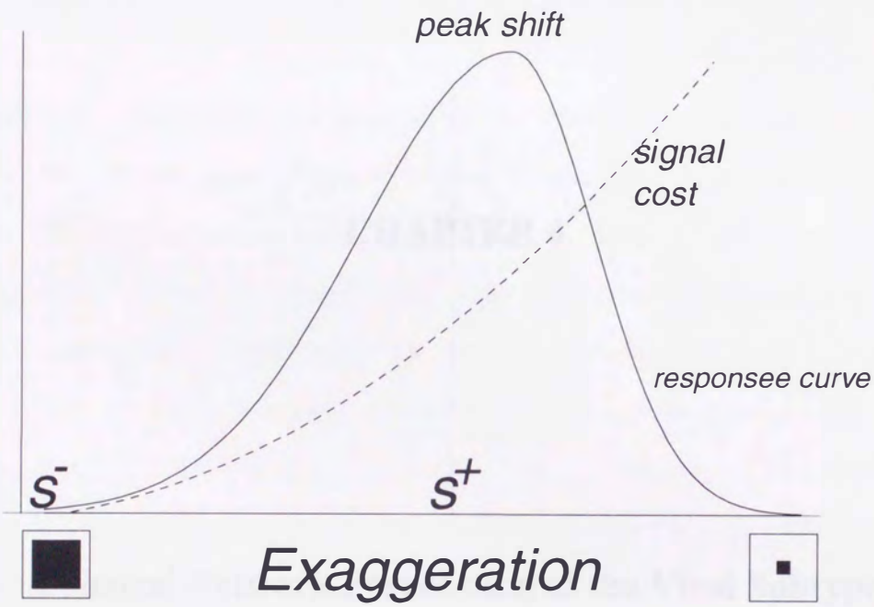
Chapter 3. Figure 4A. M. Kamo



Chapter 3. Figure 4B. M. kamo



Chapter 3. Figure 5. M. Kamo



CHAPTER 4

A Neural Network Forecasting of the Viral Subtype Dynamics: Applications to the Echovirus Epidemics.

This chapter was done in collaboration with Professor Akira Sasaki.

INTRODCTION

Echoviruses are known as a causative agent of the aseptic meningitis epidemics in Japan. Their outbreaks in Japan show seasonality: the reported number of infected hosts becomes the largest mostly in July (Figure 1A). The seasonality in outbreak is common among infectious disease, as in the widely known example of flu outbreaks.

Many epidemiological models have been studied to account for the sustained oscillation in the number of infected hosts (e.g. Nisbet & Gurney 1982, Hethcote & Yorke 1980, Schwartz & Smith 1983, Schwartz 1985). These models reveal that the introduction of seasonally varying transmission rate, for example, well explain the annual, the biennial, and the other periodicity in the outbreaks of pathogen. Schwartz & Smith (1983), for example, showed in their SEIR model that the temporal pattern of the infected density is drastically changed by the strength of seasonality.

However, these models have been focussed only on the dynamics of the single strain epidemics. In the case of echovirus epidemics, there are almost 30 serologically defined sub-types (echo1, echo2,...), and the major types change year by year. Recent nationwide outbreak occurred in 1991 in Japan with the major type echo30. In 1992, however, very few echo30 infections has been reported, and instead, echo9 caused the most of infections (Infectious Agents Surveillance Report, 1998).

Figure 1B shows the reported numbers of infections of echovirus sub-types from 1983-1986 in Japan. Although there are 33 sub-types in echoviruses, three sub-types are shown in the figure. A roughly ordered emergence by a set of sub-types is a characteristic of the echovirus outbreaks – the same subtype tends to reemerge for about every 10 years. This contrasts with the pattern of outbreaks by influenza A viruses, in which each year's nationwide outbreak is caused by a new type which is slightly deviated from the last year's major strain (antigenic drift). To understand the dynamics of those epidemics caused by a set of antigenically diverse sub-types, the extension of mathematical models is necessary. A purpose of this study is to develop an

epidemiological model dealing with the multiple strains, and to forecast the number of infected and the yearly changes in the major sub-types.

In the analysis of the real data, it is sometimes very difficult to decide whether the time series data is chaos or not because it usually contains random noises. Ellner and Turchin (1995) proposed the statistical method to estimate the Lyapunov exponents for the underlying dynamics of the time series (the positive Lyapnov exponent implies chaos). Very few studies are directed to forecast of epidemics using the empirical data. For example, using the empirical time series data, the forecasting of the measles has been done by using the SEIR model and the feed forward neural network model (Ellner et al. 1998). Even fewer studies consider the forecasting of epidemics with more than one strain (but see Fitch et al. 1997). Fitch et al. (1997) studied the forecasting of the long-term trend of human influenza viruses, and they consider the philogenical changes However, in their study, the epidemiological dynamics and the change in the number of infecteds are ignored. Hence we need a new statistical model which can predict both the sub-type change and the number of infected hosts.

The reason why the studies for the multiple sub-type model have not developed yet is in its dynamical complexity. In this paper, we propose a neural network model which can forecast the dynamics of multiple sub-type epidemics, like echoviruses. Neural networks are nonlinear model inspired by the nerve systems in animals, and they can treat systems whose inputs-outputs relationships are very complex. It is known neural networks are quite effective as a chaos detectors (e.g. Ellner & Turchin 1995, Weigend et al. 1990), and hence we expect that they might be useful for the short-term forecast for a chaotic time series. In this paper, we show that they can indeed forecast the chaotic data. Before we show that we investigate the conventional SIR model because it can be used to generate the temporal series data to train ad test the neural networks.

SIR MODEL

We use conventional SIR model with seasonality to generate the time series data of epidemics to test the performance of a predictor. The densities of the susceptible (S), the infected (I) and the recovered (R) hosts change with time as

$$\frac{dS}{dt} = -\beta SI - \mu S + \mu$$
$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I$$
$$\frac{dR}{dt} = -\mu R + \gamma I \qquad (1)$$

where β is the transmission rate, μ is the natural death rate and γ is the recovery rate. The total host density is scaled to 1. Because echovirus shows strong seasonality in morbidity, we introduce the seasonal change in transmission rate:

$$\beta = \beta_0(1 + \delta \sin(2\pi t)) \qquad (2)$$

where β_0 is the base transmission rate, and δ $(0 \le \delta \le 1)$ measures the degree of seasonality. The time is scaled in the unit of years, so the period of cycles is 1. This model shows the diverse dynamical behavior, particularly depending on the strength of the seasonality (Schwartz and Smith 1983). Assuming that the basic reproductive ratio, $R_0 = \beta_0 /(\mu + \gamma)$, is greater than 1, the endemic equilibrium is stable when δ is sufficiently small. When δ exceeds a certain threshold, two years cycle emerges. By further increasing δ, the system shows a cascade of period doubling bifurcation (2-years, 4-years, 8-years), and even more complicated series of bifurcation, depending on the parameters, towards chaos. Figure 2 illustrates the bifurcation diagram of the density of infected when δ is varied.

When this SIR model is extended to multi-strains model, the epidemics of different strains affect each other through cross-immunity. Because the densities of infected hosts are very small in realistic values of epidemiological parameters, β_0=1500 and γ=50, it is reasonable to ignore double infection in multi-strains models. We need 8 differential equations to describe the densities of each host:

$$\frac{dx_{SS}}{dt} = -\beta_1 x_{SS} x_{I\bullet} - \beta_2 x_{SS} x_{\bullet I} - \mu x_{SS} + \mu$$

$$
\begin{aligned}
\frac{dx_{IS}}{dt} &= \beta_1 x_{SS} x_{I\bullet} - \gamma_1 x_{IS} - \mu x_{IS} \\
\frac{dx_{RS}}{dt} &= \gamma_1 x_{IS} - \sigma\beta_2 x_{RS} x_{\bullet I} - \mu x_{RS} \\
\frac{dx_{SI}}{dt} &= \beta_2 x_{SS} x_{\bullet I} - \gamma_2 x_{SI} - \mu x_{SI} \\
\frac{dx_{RI}}{dt} &= \sigma\beta_2 x_{RS} x_{\bullet I} - \gamma_2 x_{RI} - \mu x_{RI} \\
\frac{dx_{SR}}{dt} &= \gamma_2 x_{SI} - \sigma\beta_1 x_{SR} x_{I\bullet} - \mu x_{SR} \\
\frac{dx_{IR}}{dt} &= \sigma\beta_1 x_{SR} x_{I\bullet} - \gamma_1 x_{IR} - \mu x_{IR} \\
\frac{dx_{RR}}{dt} &= \gamma_1 x_{IR} + \gamma_2 x_{RI} - \mu x_{RR}
\end{aligned}
\qquad (3)
$$

where x_{SS} is the density of hosts that are susceptible to both strains, x_{SI} is the density of hosts that is susceptible to the strain 1 but is infected by the strain 2 and so on. $x_{I\bullet} = x_{IS} + x_{IR}$ and $x_{\bullet I} = x_{SI} + x_{RI}$ respectively is the total number of the strain 1-infected and the strain 2-infected hosts. β_i and γ_i are the transmission rate and the recovery rate of the strain i (i=1,2). The new parameter σ stands for the degree of cross-immunity between strain 1 and strain 2. The other parameters are the same as in the Eqs. (1). Again, the seasonality is introduced as in Eq. (2): $\beta_i = \beta_{i0}(1 + \delta \sin 2\pi t)$.

Figure 3 illustrates the bifurcation diagram of two strain model for increasing δ, which is parallel to that in one strain SIR model. It shows the annual cycle when the seasonality is sufficiently small, shows 2-years, 4-years cycles, and the other cycles for intermediate strength of seasonality, and shows chaotic behavior for strong enough seasonality. However, the bifurcation diagram shifts quicker to chaos in two strains model by increasing the degree of the seasonality. Two strains SIR model shows more complicated bifurcations than that in Fig. 2 for single strain case. In Fig. 3, only the density of hosts infected by the strain 1 is plotted because we assumed symmetric parameters for both strains ($\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$), and hence the strain 2 shows the identical diagram. In these figures, we observe multiple attractor (e.g. both the annual and the biennial cycles are stable, and the trajectory is attracted to one of them independing on the initial state).

Fig. 3 also illustrates how the strength of cross immunity affects the dynamical behavior (each panel in Fig. 3 differs only in the strength of cross immunity). When the

cross immunity is very weak ($\sigma \cong 1$), two strains behave almost independently. When it is large ($\sigma \cong 0$), the competition between 2 strains become very strong. With stronger cross immunities shown in Fig. 3A and 3B, the chaotic behavior appears at a smaller seasonality parameter than in Fig. 3C for a weak cross immunity case. Both in a strong (Fig. 3A) and a weak (Fig. 3C) cross immunity cases, we observe a broad region of parameter for limit cycle attractors. On the other hand, with an intermediate strength of the cross immunity, we observe only a narrow region of parameter for one year cycle and two years cycle attractors – in the rest of region the dynamics is chaotic.

In two strains SIR mode, when the trajectory is attracted to a limit cycle with more than two years period, we expect that two strains may fluctuate with different phases. This is indeed the case: when the trajectories are attracted to two years cycles, the two strains fluctuate in perfectly opposite phase; when they are attracted to more than 3 years cycles, not only the phase also the shapes of trajectories are different. When the dynamics shows a one-year cycle, by contrast, the two strains fluctuate perfectly in phase, as expected. Unfortunately, however, the overall behavior including the detailed parameter dependence on the bifurcation structure and the phase relationship of the two strains SIR model have not yet analyzed.

The analysis of just two-strains SIR model is quite involved. If we further increase the number of strains in SIR model, mathematical analysis becomes practically impossible. Even worse from the perspective of forecasting is the sensitive parameter dependence on the dynamical behavior of multi-strain SIR model with seasonality, as shown in Figs. 2-3. Suppose that we are trying to forecast the echovirus epidemics by fitting parameters in the multi-strain SIR model. We then expect that even a small error in the estimated parameters would produce completely wrong predicted time series. To forecast the dynamical changes of the multi-strain epidemics of echoviruses or influenza viruses, we need to develop more flexible (easy to handle many strains) and fail-safe (e.g. robust against the noise included in data) models. The neural networks would be one of them.

THE PREDICTOR: NEURAL NETWORK MODEL

Neural network is a kind of a nonlinear model which is inspired by the nerve system in animals. Neural networks are considered as the connection of many neurons with certain weights. Each neuron receives n input signals of the intensity $x_1, x_2, ..., x_n$ from each connected axis with the weights of connections $w_1, w_2, ..., w_n$. A negative weight ($w_i < 0$) represents the inhibition of the neuron and a positive weight does the activation of the neuron. An activity level, or the output from the neuron j (x_j) is computed as

$$x_j = \frac{1}{1 + \exp(-\sum_{i=1}^{n} x_i w_{ij})} \qquad (4)$$

where w_{ij} is a connection weight between neuron i and neuron j and x_i is an activation level of the i-th neuron. The sum in the RHS of (4) is taken for all neurons i connected to the neuron j. The network's training is done by modifying the connection weights by a certain algorithm such as back-propagation or genetic algorithm (Bose & Liang 1996).

It is apparent that the network's ability is affected by network's architectures, such as the number of neurons in a layer or the number of layers in the network. The best number of neurons or the best number of layers for the forecasting is not decided easily. In this study, we used three-layered neural network with an input neuron, ten hidden neurons and an output neuron as a predictor for the temporal processing of data. This architecture is the same as a one-step predictor neural network used by Haykin (2000) because his networks showed the fine forecasting. The network receives an observed data at time t and given an output as a predicted data at time $t+1$. In the input layer, it has time delayed memories of order p and those are also connected to neurons in the hidden layer with some weights (Fig. 4).

Initial connection weights of the network are uniform random number s between –1 to 1. The network's training is done to minimize the squared deviation of the observed data from the predicted data. In this study, the temporal back-propagation algorithm is used following (Hykin 2000).

The observed data (say, Obs) of time t is given to the network and predicted data at time t+1 (say, Pred) is calculated by the network. Then the error is defined as

$$E(t) = \frac{1}{2}(\text{Obs}(t+1) - \text{Pred}(t+1))^2 \qquad (5)$$

The network's weights are modified to minimize this error function throughout all the training set. Since the network is a kind of function of connection weights, the error function can be differentiated by the weights. The new weight sets are given:

$$\mathbf{w}_{ij}^{new} = \mathbf{w}_{ij} - \eta \frac{\partial E(n)}{\partial \mathbf{w}_{ij}} \qquad (6)$$

where η is a learning parameter $\eta > 0$.

Before starting the prediction of the real time series data of echoviruses, the network's prediction power is tested using the data generated by the SIR model in Eqs. 1. For network's training, data are rescaled between 0.1 to 0.9 because the densities of infected hosts generated by the SIR model is sufficiently small and is not suitable for the network outputs. The parameters for the network are in Table 1.

Of course, the network's prediction power is affected by these parameters, especially the number of the memories in the input layer and the number of data points for training are particularly important. We sampled 25 data point per year, and adopted 100 memories of the input layer is corresponding to the 4 years. It is possible to adopt more memories in the input layer and more numbers of data points for training, but we did not because those are "enough" for the network's training. We can expect that there should be some optimal architecture for the network, however we could not find the good references for that.

Observed data are given to the network for the first 100 data points (8 years), and network's outputs (prediction) is feeded back to the network's input neuron for another 200 data points. Hence the network's predictions after 400 data points are done by their prediction data sets.

RESULT

Figure 5 illustrates the result of the prediction. Fig. 5A shows that the observed data generated by the one strain SIR model and Fig. 5B shows the results of a prediction of the network. In this figure, the seasonality intensity (δ) is 0.05 and the density shows the two years cycle. The networks were trained to fit to the generated data using the first 8 years. During the training period, the network's predicted data at time t+1 is used as the network's input of time t in the next step. The correlation between generated data and the predicted data equals 1 for, at least, another 60 years. The network's prediction powers were tested using the data sets which shows one year cycle (when δ=0.01). The correlation between generated data and the predicted data were exactly 1 for very long time. This is very trivial result because the predictions for those simple data sets can be done by another method, such as some statistical method using autocorrelation, hence we do not have to use the neural networks. However these are the good sample to know that the network also can be a good predictor.

When the dynamics shows the chaotic behavior, can the network still be a good predictor? The chaotic data were generated using the SIR model using the larger seasonality (δ=0.1 see Fig.2) in which the cycle has two years periods, but the peak amplitudes vary irregularly. The networks were trained with the same architectural parameters above. Fig. 6A shows the generated data and the predicted data. The first 8 years were used as the training sets, therefore network's prediction is perfectly overlapping to the generated data. This means that the network's training was successful. After 8 years, the network's prediction started. Fig.6B shows the correlation between generated data and the predicted data. When the correlation is calculated, the training set was excluded. We used the first four years data sets as the default, and the correlation was calculated till 80 years. The Fig.6 indicated that the correlation was quite high for the first few years, but soon went down. This implies that the network's long term prediction was not good although, but the short-term prediction was good.

Sensitivity to the noise.

Next we investigated the network's prediction power using noisy data. Noise in this case means the sampling noise. After the densities of the infected hosts were generated by using the SIR model, normal random numbers (m=0, sd = 0.1) were added to the data. In this case, a parameter for the network was changed. The changed parameter is the number of back propagation cycles for training. The number of back propagation cycles was 1000 rather than 100000. The reason for this will be mentioned later. When the density data was generated, the strength of the seasonality (δ) was 0.05 which is in the parameter region where two years cycles are mostly obtained (Fig. 2). Figure 7 shows the correlation between generated data and the predicted data. When no noise is added to, the correlation was 1 until at least 80 years. This time the correlation decreases as the time for the prediction went by.

We may expect that for the network's learning, when more back-propagation cycles are obtained, the prediction power of the network should be better. Indeed, when there are no sampling noise in the data, the more is the number of back-propagation cycles, the better is the performance of the trained network. However, when the sampling noise involved, too much learning might amplify learning error because of the overfitting to the noisy component of the data. To examine this idea, the network was trained in various numbers of back-propagation cycles. Again the strength of the seasonality is 0.05, hence the densities showed two years cycles. The result is shown in Figure 8. In Fig. 8, the correlation between predicted data and the generated data are shown. The correlation within the training set is shown in a dashed line and the correlation using the data after the training period is shows in a solid line. In the case of solid line, the correlation was computed using the 68 years data points. Since the networks were trained to fit the prediction to training tasks, if more number of back-propagation cycles were obtained, the correlation should go up (dashed line). On the other hand, the correlation computed using the predicted sets went down when the number of back propagation cycles becomes large. We can expect that this phenomenon is caused by overlearning. Usually the densities of the infected hosts by some viruses

contain many noises, probably brought by randomly varying environment condition, such as different mean temperature or mean humidity of the year. The figure warns us that we always have to be careful for avoiding the over learning of the network.

In this study, we just concerned sampling noise in the data sets. However, we can also concern the random fluctuation in the transmission rate itself. In this case, we can expect that more complicated dynamics appear. Because they show two years cycles or three years cycles depending on the initial condition even when the strength of the seasonality is the same (see Fig. 2 and Fig. 3), the dynamics which is on the stable orbit of two years cycle will sometime move onto the orbit of the stable three years cycle. We have to examine whether the networks can still predict such a time series data, but the investigation for the orbit changes has not bee done yet. This should be our future study.

Two strains predictor

The modification of the networks for the two strains predictor is very easy, we can just add an input neuron to the input layer. The networks had the same parameters sets as shown in Table 1, but the number of input neurons and the output neurons are doubled. The training set for the neural networks were generated using two strains SIR model defined in Eqs. 3. The bifurcation diagram of two strains SIR model is shown in Fig. 3. In fig. 3, we only investigated the dependencies of δ, the dynamics is also affected by the strength of the cross immunity which is indicated as σ in Eqs. 3. The bifurcation diagram changes according to the intensity of the cross immunity (Fig. 3). However, the effect of the cross immunity to the bifurcation structure is not yet fully analyzed. In this study, we fixed the σ to 0.9. because with that amount of σ we can obtain more complicated seasonal fluctuations when we compared the smaller σ (see Fig. 3A and Fig. 3C). With that amount of σ, both strains affect a very little and behave almost independently.

First we generated the densities of infected hosts using δ=0.01. In this case, the densities show an annual cycle (see Fig. 3C). When computing the correlation, the

training sets were excluded. The first four years correlation were calculated as default and the other prediction sets were added to the data set until 80 years from the beginning. The correlation coefficient for both strains (r_1, r_2) was 1 until at least 80 years for prediction. This implies that the network's prediction power is not worse even when the network is modified multi strains predictor.

THE FORCASTING THE ECHOVIRUSES MORBIDITY

Now we try to forecast the real data. The number of isolations is obtained from Jpn. J. Med. Sci. Biol. (Suppl. 1983-1976). Theses are monthly data and the real number is shown in Figure 1. Figure 9 shows the rescaled between 0.1-0.9 for the network training and is shown in logarithmic scale. For the network training, the first 7 years data are used and the observed data of the last 7 years and forecasted data are compared. Evaluation is done using the correlation between observed data and the forecasted data. The network's parameters in this case are in Table 2.

The result for this forecasting is shown in Figure 9. Fig. 9A shows the result of the forecasting and B shows the correlation. For the correlation calculation, the training set is excluded. Exact amount is not predicted, but the global tendency, such as changes of the maximum densities of the years, was predicted well.

DISCUSSION

In this paper we introduce a new method of forecasting the epidemiological time series data, such as the monthly change in the densities of hosts infected by echoviruses or influenza viruses. Diversity of pathogen strains complicates the task: there are at least 30 sub types caused the outbreaks of echoviruses during last two decades; antigenically different mutants come up every year in the flu outbreaks. For forecasting such epidemiological outbreaks caused by multiple causative agents, we need to have more flexible predictors than what has been developed so far for a single strain epidemic. Previous widely known theoretical models dealing with the population dynamics of infectious diseases (e.g. Anderson & May 1979) have focussed on a single strain dynamics, and the effect of the multiple strains have been ignored. Furthermore, quite a few studies have been directed to the forecasting of the epidemiological dynamics (Ellner & Turchin 1995). For practical purposes like the exploitation of the effective vaccination and the decision on the amount of vaccine stock in a year, the prediction on the morbidity and the major subtype in the next year is critically important. The reason why the effect of multiple strains has been ignored is in its difficulty in coping with the enormous complexity in the dynamics, but without the effort of understanding the multiple strain dynamics, the prediction is of no use in many infectious diseases.

The seasonality is clearly important in echoviruses and influenza virus epidemics. The morbidity attains the peak in summer in echovirus infections (see Fig.1) and in winter in influenza virus infections. Theoretical studies on the epidemiological dynamics have revealed that the introducing of even a small amount of seasonal variation in the epidemiological parameter drastically change the temporal pattern (the annual, the biennial, and the other subharmonics) of outbreaks (e.g. Nisbet & Gurney 1982, Yorke et al. 1979). Shcwartz & Smith (1983) analyzed the bifurcation structure of the SIR model with seasonally varying transmission rate. They found that the population shows a cascade of period doubling bifurcation by increasing the strength of seasonality. However, the bifurcation structure of the multi-strain SIR model with

seasonality has not been studied yet.

When we extend the SIR model to multi-strain model, we have to consider the effect of cross immunity. Most theoretical models with multiple strains have assumed either perfect cross-immunity (once a host acquires the immunity to a certain strain, it will never be infected by any other strains) or no cross-immunity (epidemiological processes of the strains are mutually independent). However, because subtypes are antigenically different (by definition), a host infected by a subtype may be reinfected by another. The secondary infection rate becomes weaker than the primary one, the extent of which is determined by the degree of cross immunity, i.e., by the serological distance between subtypes. The behavior of the multi-strain epidemiological dynamics critically depends on the coefficients of cross immunity defined for each pair of subtypes (Fig.3).

In this study, we used the conventional SIR model with seasonality as the data generator, because it is widely accepted and can generate a wide variety of trajectories from a simple annual cycle to a chaotic one, according to the strength of seasonality. Before we introduce the predictor, we investigate the way the bifurcation occurs in single strain system by numerical simulations (summarized in Fig. 2) and in the two strains SIR system (Fig. 3). We found that the pattern of the bifurcation differs significantly when the degree of with different cross immunity between strains is changed. A possible approach for the forecasting is to estimate the crucial epidemiological parameters, like the strength of seasonality and the cross immunity, from the observed time series of the dynamics. However, because the number of parameters that are to be estimated increases rapidly as the number of strains increases, this is not a simple task.

Our approach in this paper is to establish a new method which can be easily extended to the multiple strains model. We adopted the neural networks model for that. The neural network is a non-linear model inspired by nerve systems in the neuron and they have the ability of learning, pattern recognition and generalization. The merit of using the neural network is its handiness to cope with complicated systems which are driven dynamical variables and may be disturbed by many kinds of noises.

We first tried the forecasting using the data generated by the single strain SIR model. When we introduce only a small amount of seasonality, the networks forecasting performs well: the correlation between generated data and the predicted stayed nearly 1 for a long time. Even when we introduce a larger amount of seasonality, e.g. δ=0.1 (see the bifurcation diagram, Fig.2), the network could forecast the densities of infected, although the precise forecasting is restricted only for near futures. This prediction power was inherited when the network was extended to the two strains predictor as long as the cross immunity is weak.

When the network was trained with the data sets which were mixed with random sampling noise, the correlation between generated data and the predicted data remains high (Fig. 7). However, as shown in Fig. 8, the prediction become worse when the network was trained too much, because the network loses its robustness in prediction by fitting to the noisy component of the data set. If there is no noise, the network performs better successful when it is trained for longer time; but with noise, an intermediate length of training is the best (Fig. 8).

After we checked the prediction power of the network, we tried to forecast the densities of echoviruses infections from the data of the reported cases. Because such data should include noises, the networks learning should be stopped, to avoid over-learning, at an appropriate number of back propagation cycles. The estimation for the accuracy of the prediction was done by using the correlation between observed and the predicted data. We obtained a high correlation between the reported and the predicted number of cases.

In this study, the networks architecture or parameters were decided by trial and error, i.e. by looking for the best architecture that gives the highest correlation between predicted and observed data. The network's architecture is particularly important when we deal with the data containing the random noise, because we have to avoid the over-training. However, we have not yet had a theoretical understanding for the optimal architecture or the method of avoiding the over-learning. The optimal design of the network as a predictor is open to future studies.

REFERENCES

Anderson, R. M. & May, R. M., 1979. Population biology of infectious diseases: part I. *Nature Lond.* 280: 361-367

Bose, N. K. Bose and Liang, P. 1996. *Neural Network Fundamentals with Graphs, Algorithms, and Applications.* New York: McGraw-Hill Inc.

Ellner, S. & Turchin, P., 1995. Chaos in a noisy world: new methods and evidence from time-series analysis. *Am. Nat.* 145: 343-375

Ellner, S. P., Bailey, B. A., Bobashev, G. V., Gallant, A. R., Grenfell, B. T. & Nychka, D. W. 1998. Noise and Nonlinearrity in Measles Epidemics: Combining mechanistic and statistical approaches to population modeling. *Am. Nat.* 151: 425-440

Fitch, W. M., Bush, R. M., Bender. C. A. & Cox, N. J. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. USA*. 94: 7712-7718

Hethcote, H. W. & Yorke, J. A. 1980. *Gonorrhea Transmission Dynamics and Control.* New York: Springer-Verlag.

Hykin, S. (2000). *Neural Networks.* Ontario: McMaster University.

Infectious Agents Surveillance Report, 1984-1987. *Jpn. J. Med. Sci. Biol. Supl.* Vol. 37-50

Nisbet, R. M. & Gurney, W. S. 1982. *Modelling Fluctuating Populations.* New York: John Willey & Sons

Schwartz, I. & Smith, H. 1983. Infinite subharmonics bifurcation in an SEIR epidemic model. *J. Math. Bol.* 18: 233-253

Schwartz, I. B. 1985. Multiple stable recurrent outbreaks and predictability in seasonality forced nonlinear epidemics models. *J. Math. Biol.* 21: 347-361

Weigend, A. S., Huberman, B. A., Rumelhart, D. E. 1990 Predicting the future: a connectionist approach. *International Journal of Neural Systems.* 3: 193-209

Yorke, J. A., Nathanson, N., Pianigiani, G. & Martin, J. 1979. Seasonality and the

requirements for perpetuation and eradication of viruses in populations. *Am. J. Epidemol.* 109: 103-123

Figure Legends

Figure 1 Figure 1A depicts the reported number of infected people by echoviruses. The number is reported monthly. The dots show the number in July. The clear seasonality is seen in every year. Figure B shows the three examples. The number of echo9 is shown in gray line, that of ehoc18 is shown in solid line and that of echo30 is shown in dashed line. They often show sub-type changes.

Figure 2 The bifurcation diagram of single strain SIR model. Horizontal axis shows the strength of the seasonality and the vertical axis shows the densities of the infected hosts. The number n (n<10) in the figure shows the n years cycle. More than ten years cycle are shown in black dots. In each strength of the seasonality, 100 simulations with different initial condition were done. The simulation was continued until the effect of initial condition was excluded. The densities in July (when the infection rate becomes the maximum) in every year are plotted. The parameters used in the simulation were β=1500, γ=50, μ=0.01. Those parameters were almost the same as in Schwartz & Smith (1983).

Figure 3 The bifurcation diagram of two strains SIR model. The parameters are the same in Fig. 2 and we put a symmetric parameters for strains ($\beta_1 = \beta_2$=1500, $\gamma_1 = \gamma_2$=50). The number n (n<10) in the figure shows the n years cycle. More than ten years cycle are shown in black dots. Therefore, the bifurcation diagram of one strain is enough, because if we simulate in various initial conditions, the diagrams of both strains are the same. A indicate the diagram when the strength of the cross immunity (σ) equals 0.1. B shows the diagram when σ is 0.6 and C shows that when σ is 0.9. There are more atractors than in Fig. 2.

Figure 4 An image of the neural network in this study. Every cell between layers is fully connected by some weights. The network's training is done by modify the weight set appropriately. Network was input the time t data and predict the time t+1

data.

Figure 5 The comparison between generated data and the predicted data. The data were generated using the SIR model with seasonality. The strength of the seasonality in this figure was 0.05 (see Fig.2). The first 8 years data are the training set, after that year, the neural network kept forecasting only using its prediction, i.e. the observed data was not give to the network during the prediction period. We continued the simulation until 80 years including training set. The correlation between observed data and predicted data (training set were excluded) was exactly 1.

Figure 6 The comparison between generated data and predicted data. In this case, δ is 0.1 and shows the chaotic behavior. A indicate generated data (solid line) and the predicted (dashed line). B shows the correlation. After the training is over, the four years correlation was calculated as default. Then by adding both data one by one, the correlation was recalculated. The correlation is high for the first couple of years.

Figure 7 Correlation between the generated data which was contaminated by random noise and the prediction for those. In this case, the seasonality strength was 0.05. It shows the two years cycle in this simulation. If the data is not contaminated, the correlation was exactly 1 for long time (more than 50 years). However, in this case, the noises are added, therefore the high correlation was not maintained for long time.

Figure 8 Correlations dependency on the learning period. Horizontal axis shows the number of back propagation cycles and the vertical axis shows the correlation. The data sets include the training tasks and the predicted sets 80 years in total. The first 8 years are the training set and the last 72 years are the predicted set. The dashed line shows the correlation within the training set; therefore, when the more back propagation cycles are obtained, it become high. Solid line was the correlation within the predicted set. Not like in the training set, when the more number of back propagation cycles are

obtained, the correlation become small.

Figure 9 The prediction in real numbers. The training set is the first half and the last half was predicted by the trained network. A shows the result of the prediction shown in log scale. B shows the correlation between them

Table 1.

The parameters of the network for forecasting the data generated by SIR model.

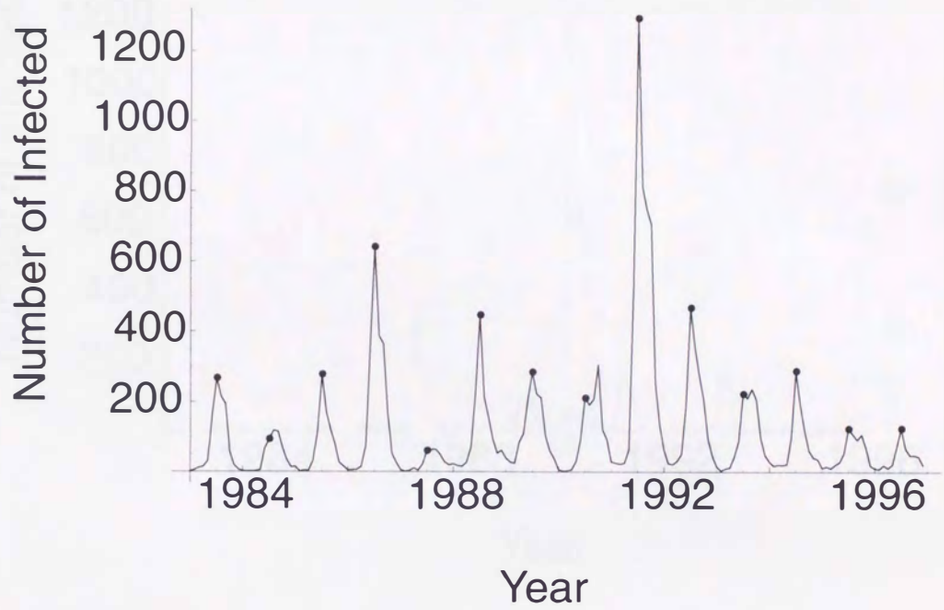
Number of input neuron	1
Number of memories	100 data points (4 years)
Number of neurons in the hidden layer	10
The back-propagation cycles for training	100000
Sampling rate of the observed data	25 points per year
The number of training set	200 data points (8 years)

Table 2

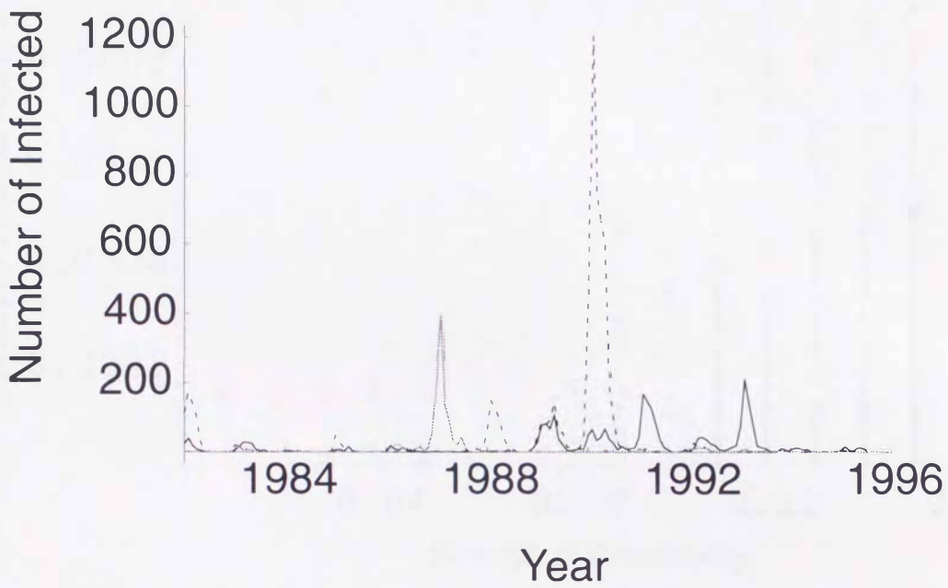
Parameters of the network for forecasting the empirical data of the echoviruses

Number of input neuron	1
Number of memories	42 months
Number of neurons in the hidden layer	10
The back-propagation cycles for training	5000
Sampling rate of the data	Monthly
The number of training set	200 data points (8 years)

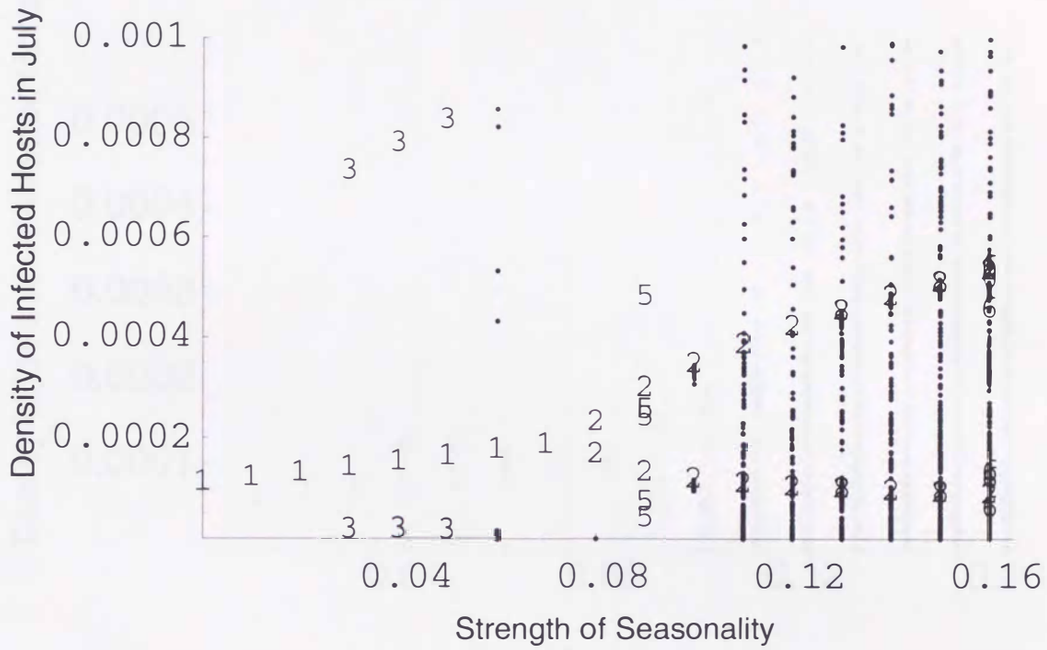
Chapter 4. Figure 1A. M. Kamo



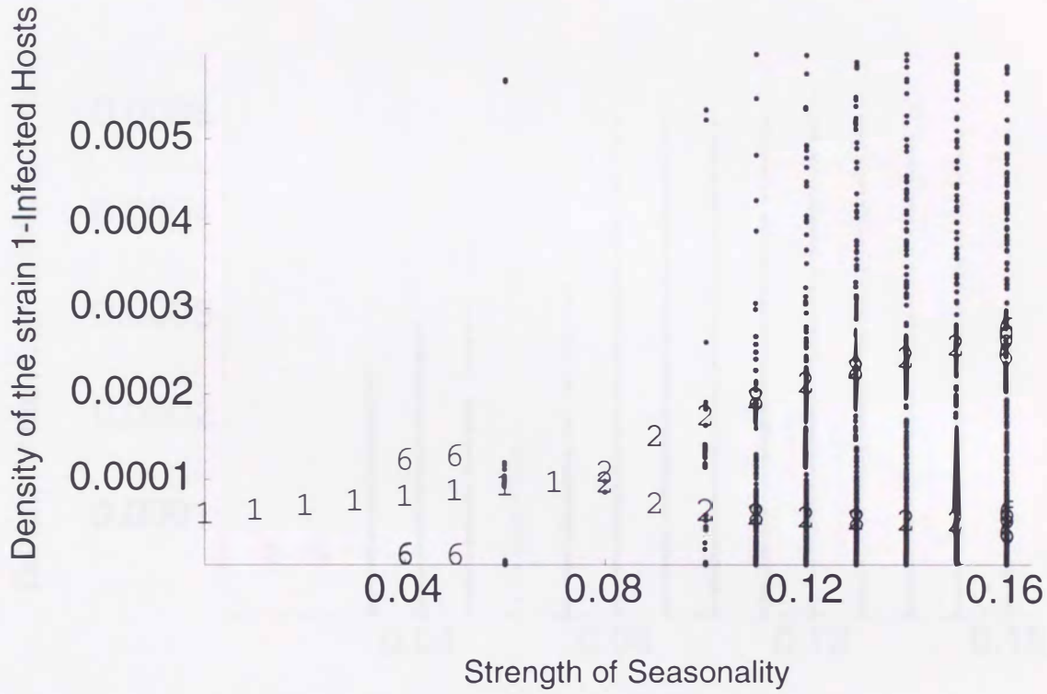
Chapter 4. Figure 1B. M. Kamo



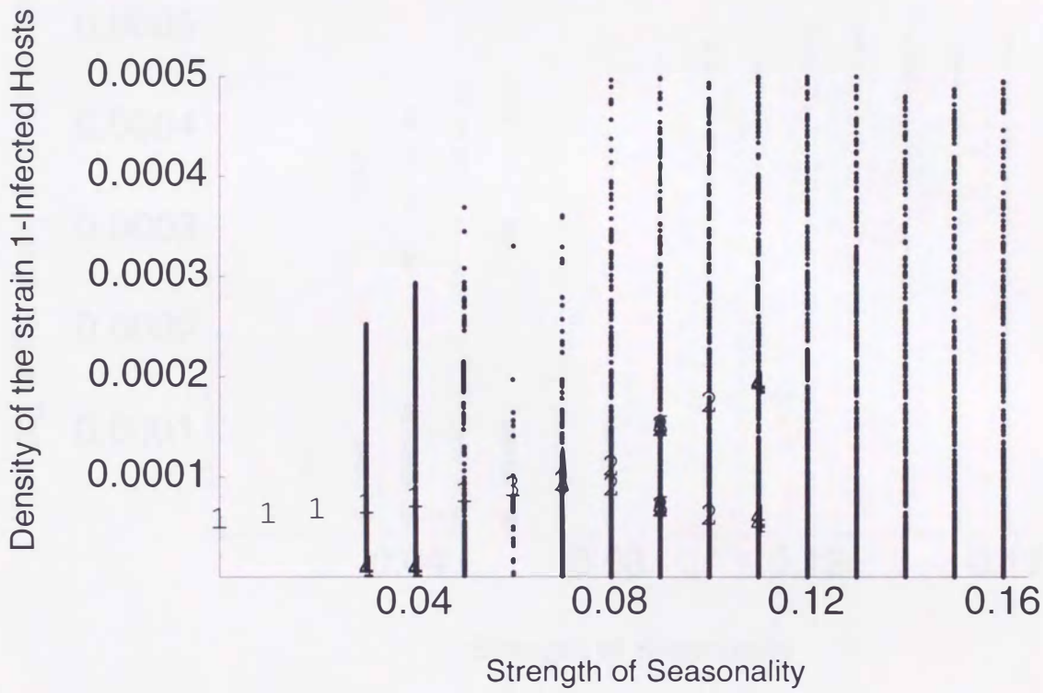
Chapter 4. Figure 2. M. Kamo



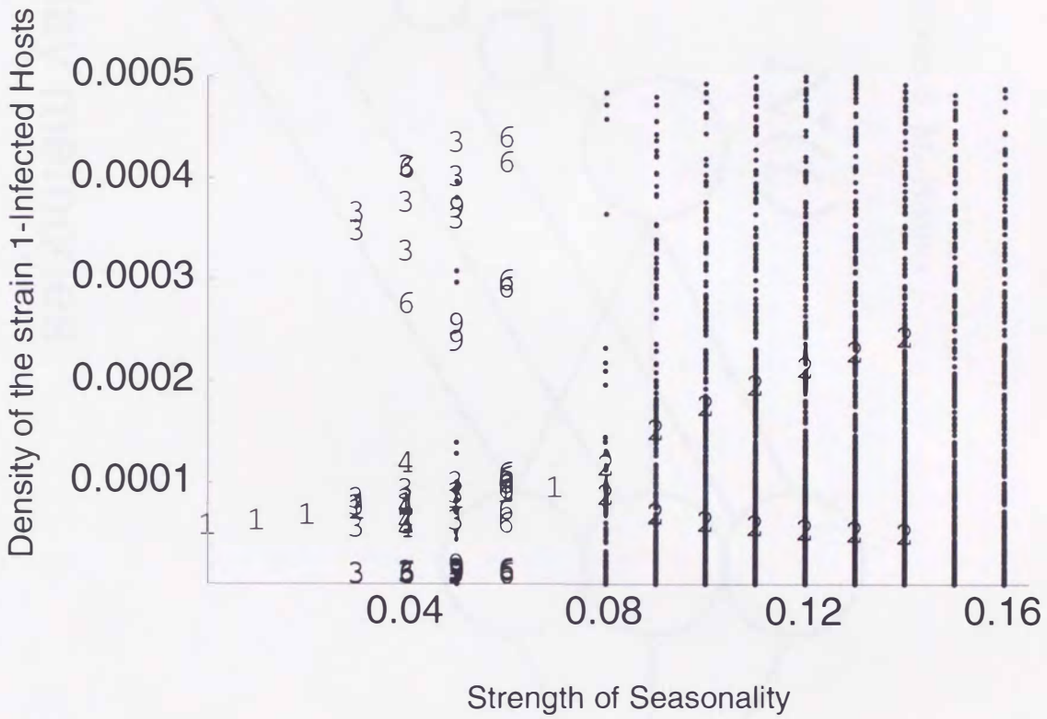
Chapter 4. Figure 3A. M. Kamo



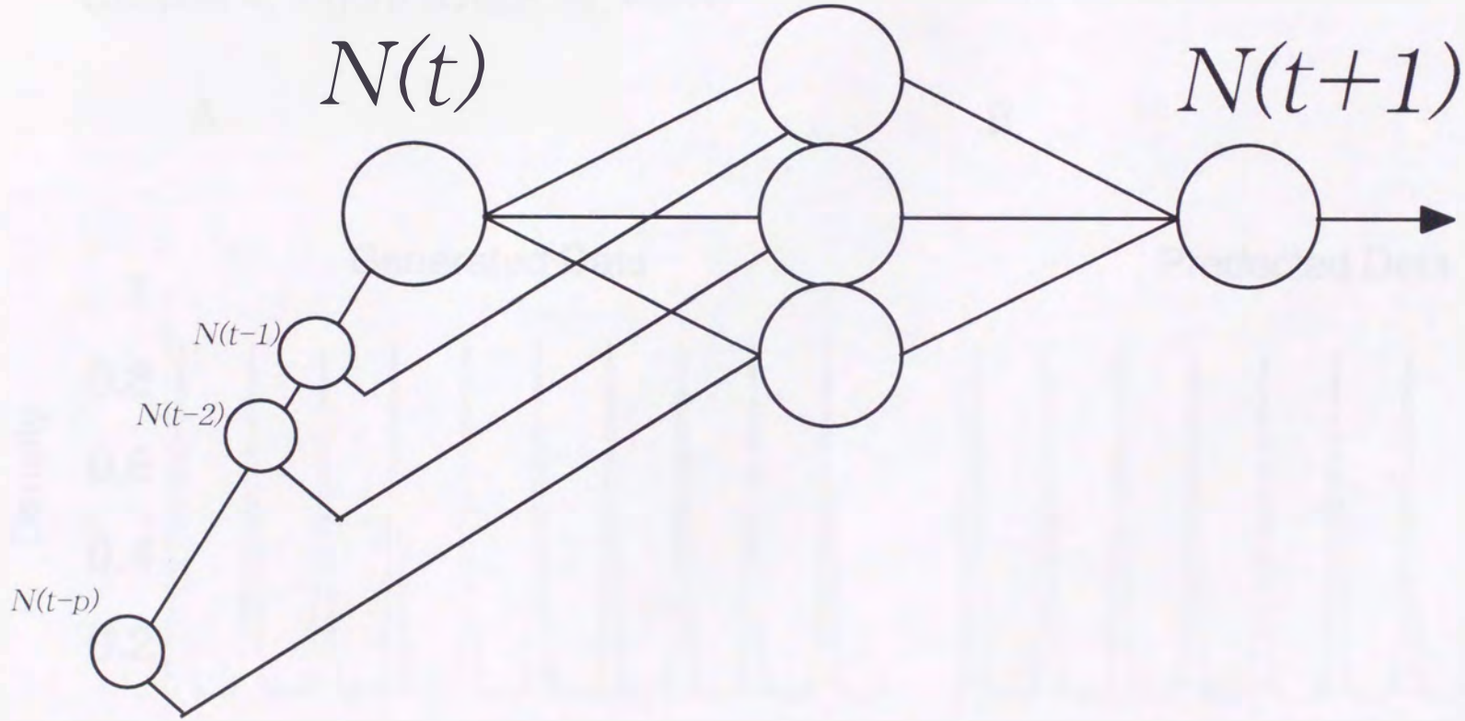
Chapter 4. Figure 3B. M. Kamo



Chapter 4. Figure 3C. M. Kamo

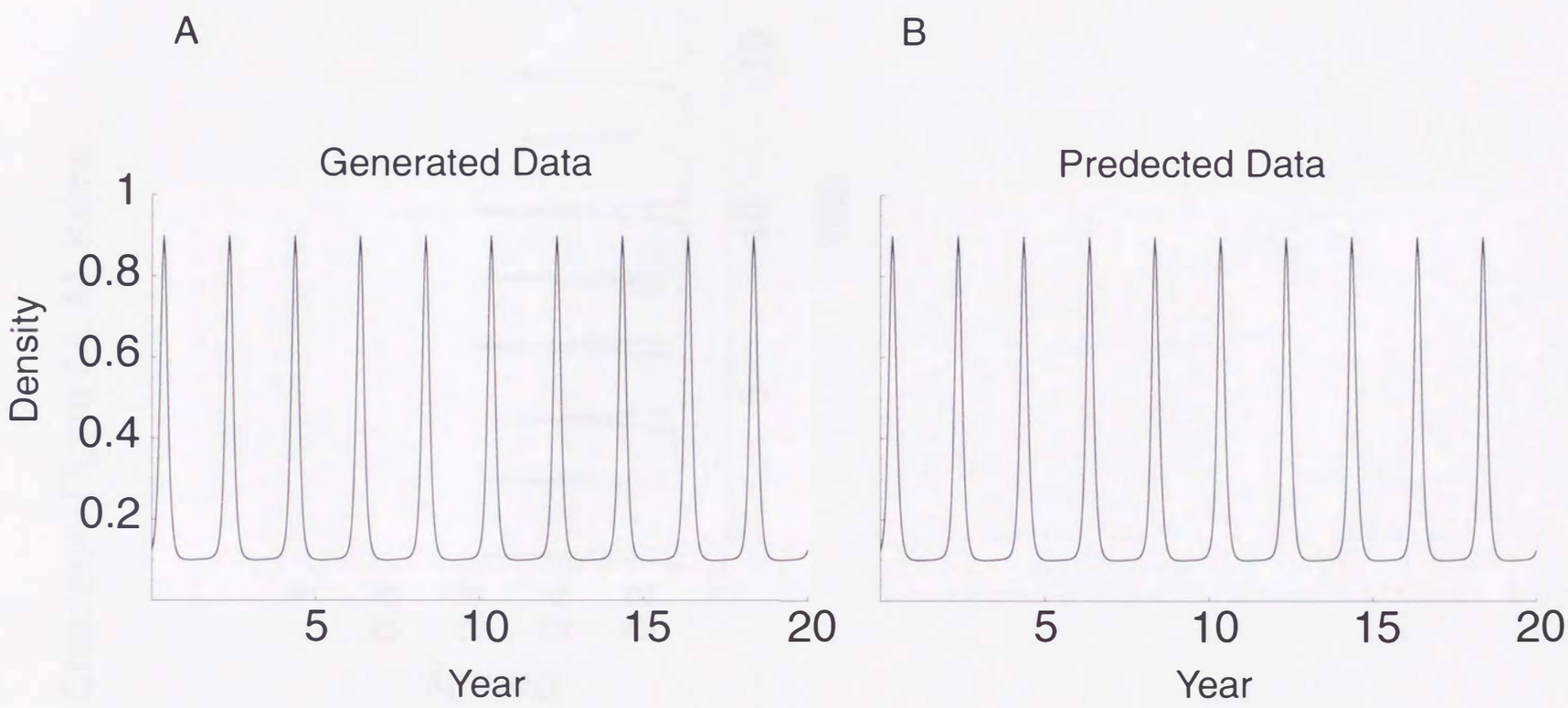


Chapter 4. Figure 4. M. Kamo

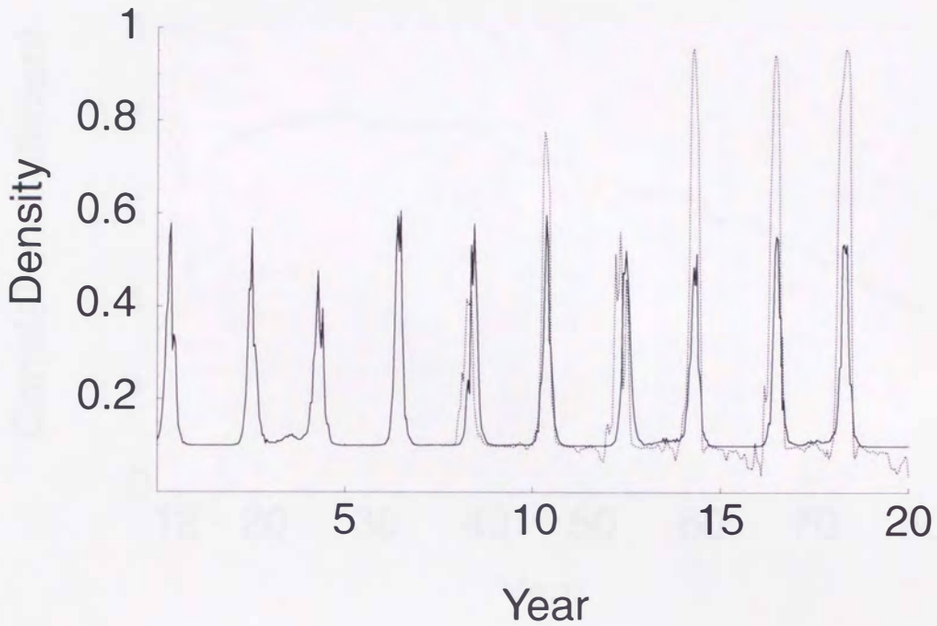


Time dalay memories

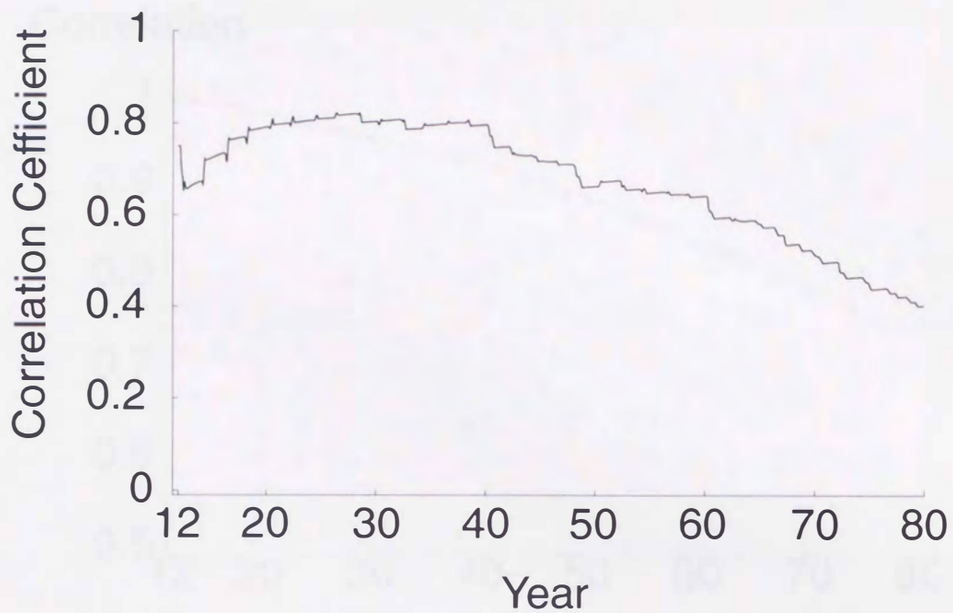
Chapter 4. Figure 5 A&B. M. Kamo



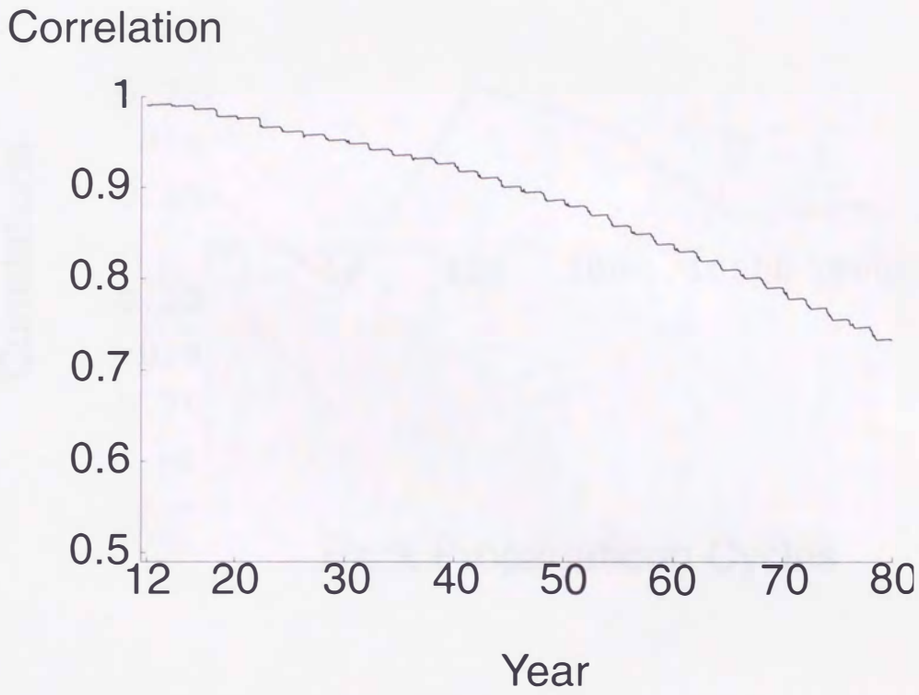
Chapter 4. Figure 6A. M. Kamo



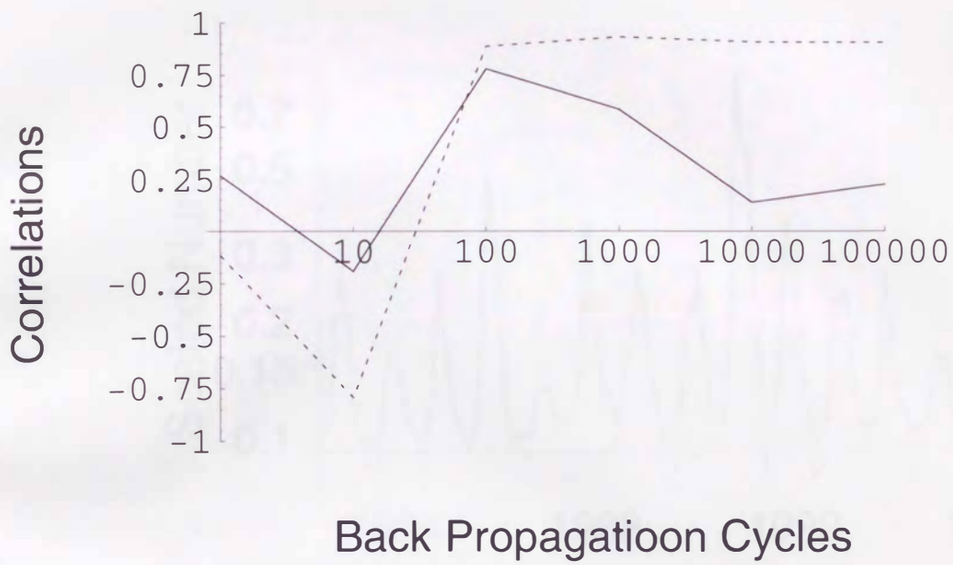
Chapter 4. Figure 6B. M. Kamo



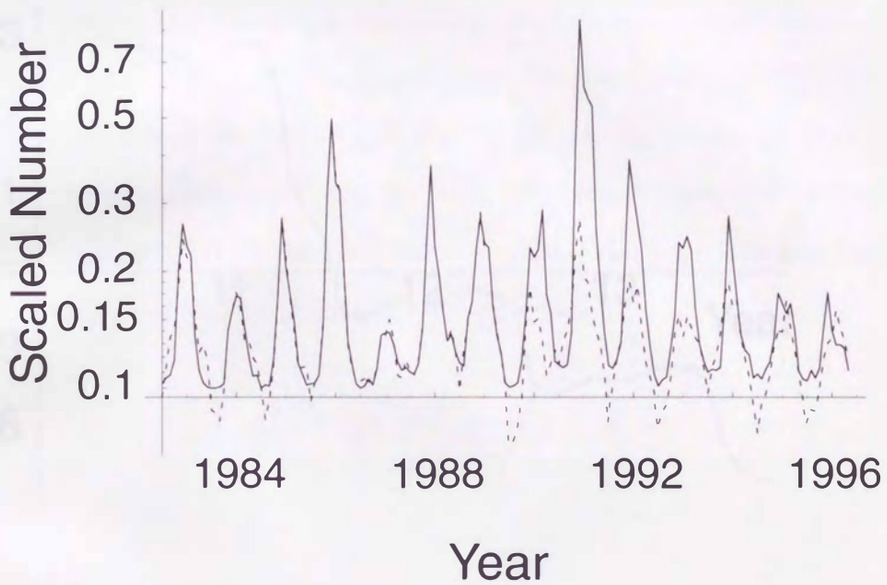
Chapter 4. Figure 7. M. Kamo



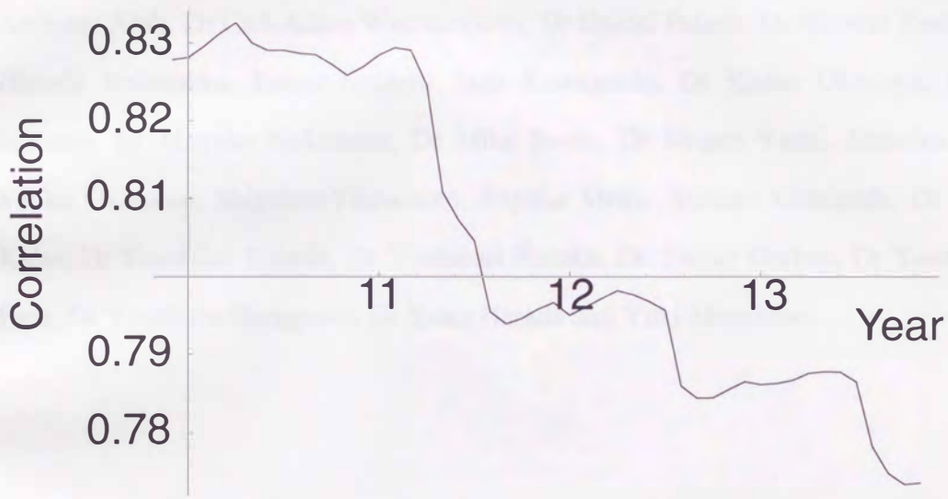
Chapter 4. Figure 8. M. Kamo



Chapter 4. Figure 9A. M. Kamo

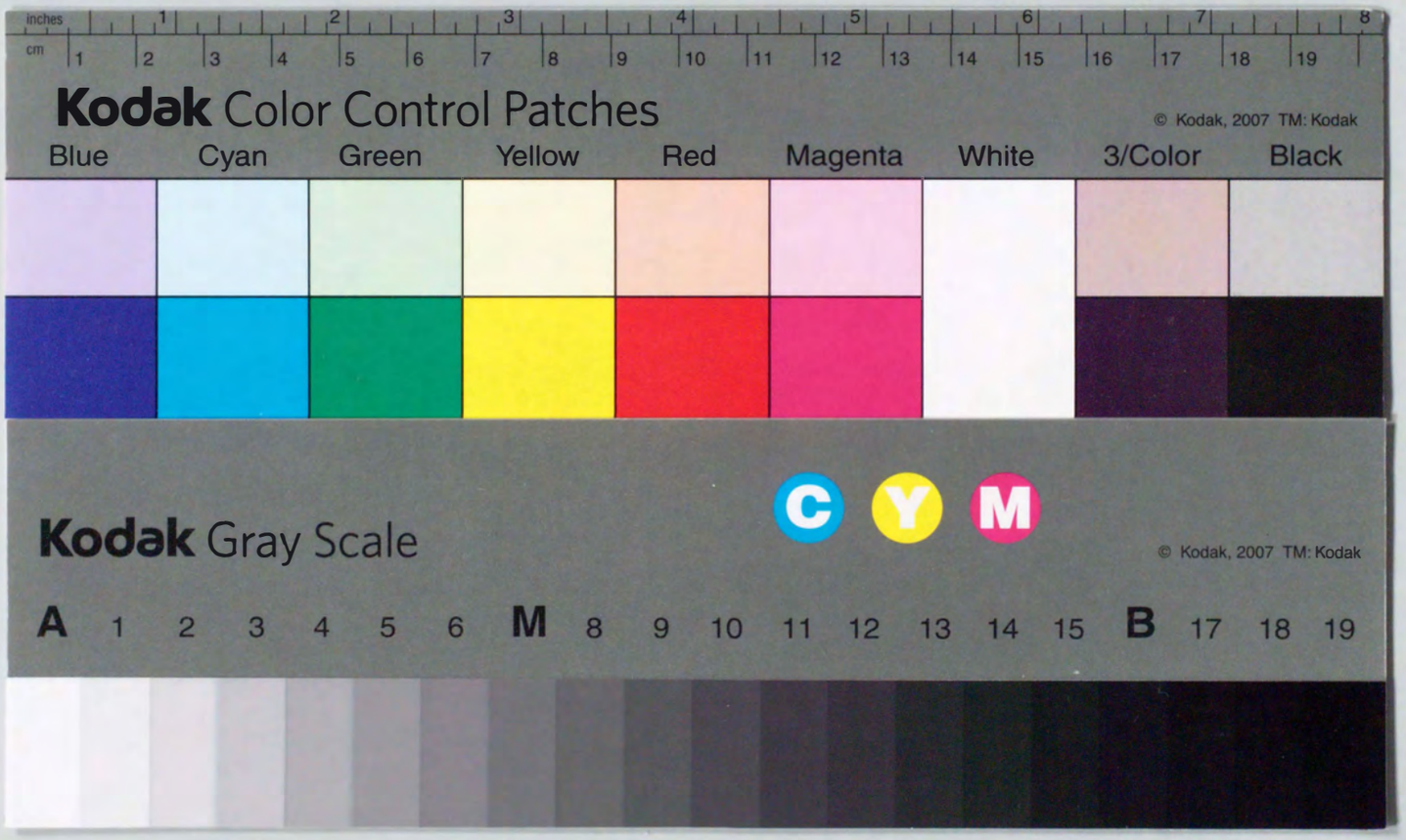


Chapter 4. Figure 9B. M. Kamo



ACKNOWLEDGEMENTS

I sincerely thank Professor Yoh Iwasa, Professor Hiroyuki Matsuda, Professor Akira Sasaki and Professor Magnus Enquist with great respects. I also thank following people for their helpful comments, technical supports and strict criticisms: Akiko Satake, Dr Anthony Arak, Dr Carl-Adam Watchmeister, Dr Daniel Falush, Dr Hiromu Yoshida, Dr Hiroshi Hakoyama, Isamu Kajitani, Isao Kawaguchi, Dr Kazuo Okanoya, Masami Isomura, Dr Mayuko Nakamaru, Dr Mike Boots, Dr Shigeo Yachi, Shusaku Tohya, Shoko Okamoto, ShigehiroYamamoto, Soyoka Muko, Stefano Ghirlanda, Dr Takuya Kubo, Dr Yasuhiko Takeda, Dr Yoshinari Tanaka, Dr Tomas Herben, Dr Young-Seuk Park, Dr Yoshihiro Haraguchi, Dr Yuko Harada and Yuki Morimoto.


inches
cm
Kodak Color Control Patches
© Kodak, 2007 TM: Kodak
Blue
Cyan
Green
Yellow
Red
Magenta
White
3/Color
Black
Kodak Gray Scale
C
Y
M
© Kodak, 2007 TM: Kodak
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19