Neural Networks, Vol.21, pp.1224-1231, 2008

Developmental Learning of Complex Syntactical Song in the Bengalese Finch: A Neural Network Model

YUICHI YAMASHITA [1], MIKI TAKAHASI [2], TETSU OKUMURA [1], MAKI IKEBUCHI [2], HIROKO YAMADA [2], MADOKA SUZUKI [2], KAZUO OKANOYA [2], JUN TANI [1]

[1] *Lab. for Behavior and Dynamic Cognition*

[2] *Lab. for Biolinguistics*

Brain Science Institute, RIKEN

2-1 Hirosawa, Wako-shi, Saitama, 3510198 Japan

Corresponding author

Yuichi Yamashita

Brain Science Institute, RIKEN

2-1 Hirosawa, Wako-shi, Saitama, 3510198 Japan

E-mail: yamay@brain.riken.jp

Tel: +81-48-467-1111 (ext. 7415)

Fax: +81-48-467-7248

Abstract

We developed a neural network model for studying neural mechanisms underlying complex syntactical songs of the Bengalese finch, which result from interactions between sensori-motor nuclei, the nucleus HVC (HVC) and the nucleus interfacialis (NIf). Results of simulations are tested by comparison with the song development of real young birds learning the same songs from their fathers.

The model shows that complex syntactical songs can be reproduced from the simple interaction between the deterministic dynamics of a recurrent neural network and random noise. Features of the learning process in the simulations show similar trends to those observed in empirical data on the song development of real birds. These observations suggest that the temporal note sequences of songs take the form of a dynamical process involving recurrent connections in the network of the HVC, as opposed to feedforward activities, the mechanism proposed in the previous model.

Key words: recurrent neural network, noise, birdsong, development, HVC, NIf, Zebra finch

List of symbols:

w_{ij}: weight value from the jth unit to the ith unit

$u_{i,t}$: internal state of the i th unit at time t

$x_{i,t}$: neural state of the i th unit at time t

$y_{i,t}$: activation of the i th unit at time t

$y^*_{i,t}$: desired activation value of output units at time t

$e_{i,t,}$: error between desired value and actual value of output activation at time t

E : learning error

T : length of note sequences

O : set of indexes corresponding to the output units

N : total number of units

α: learning rate

θ_i : threshold of the ith unit

G : noise added to the internal value of non-output units

F_{max}: transformed value of the unit with the highest activation in the winner-take-all computation

F_{min}: transformed value of all other output units in the winner-take-all computation

A : component notes of a particular song

D : a particular song syntax with a probabilistic distribution of strings

$P_D(x)$: occurrence probability of string x under distribution D

1. Introduction

Because of its similarity to human language in being a learned complex sequential behavior, birdsong, has become a widely studied topic in neuroscience. The Bengalese finch in particular learns highly complex songs that have syntactical structure, providing researchers with a good biological model for studying this phenomenon.

Figure 1 shows a typical sound spectrogram of the song note sequences of the Bengalese finch. The song consists of several varieties of "notes", the smallest units of a birdsong. Each note can be identified as a discrete element on a sound spectrogram and is denoted by a letter of the alphabet, for example "a", "b" or "c".

Note-to-note transitions follow rules. However, the transitions are not determined by the preceding note output alone, but are dependent on past sequences. This indicates that the Bengalese finch's song sequence has a hidden state in the sense that the next transition state cannot be uniquely determined by the output note.

Note-to-note transition rules of the Bengalese finch's song can be described using a finite state automaton (Honda and Okanoya 1999). Normally the automaton describing a Bengalese finch's song has probabilistic branching and recursive connections. A series of notes without branching constitutes what is referred to as a "chunk", and sequences of chunks generate diverse "motifs". Owing to the recursive structure of the automaton describing their songs, the Bengalese finch is considered to generate an almost infinite number of different motifs. The complexity of this song structure is in contrast to the linearity of the songs produced by the Zebra finch, a bird which is nonetheless a close relative of the Bengalese finch (Zann 1996).

The acquisition and production of songs is made possible by a group of discrete brain nuclei and their connecting pathways, referred to as the song system (Fig. 2) (Nottebohm 2005). Within the song system the nucleus HVC (HVC), a premotor nucleus, plays a key part in generating temporal patterns of songs. In lesion studies of the Canary, a lesion on the HVC severely disturbs the temporal structure of songs (Nottebohm et al. 1976). In electrophysiological studies of the Zebra finch, the activation pattern of each HVC neuron is highly context-dependent and corresponds to a particular moment in a song (Fee et al. 2004). Moreover HVC stimulation by a microelectrode of a singing bird causes an interruption in temporal patterns of the birdsong, whereas stimulation of the robust nucleus of arcopallium (RA), a downstream motor nucleus, affects only a particular note at the time of the stimulation (Vu et al. 1994). These facts strongly suggest that the HVC is a temporal pattern generator for song note sequences.

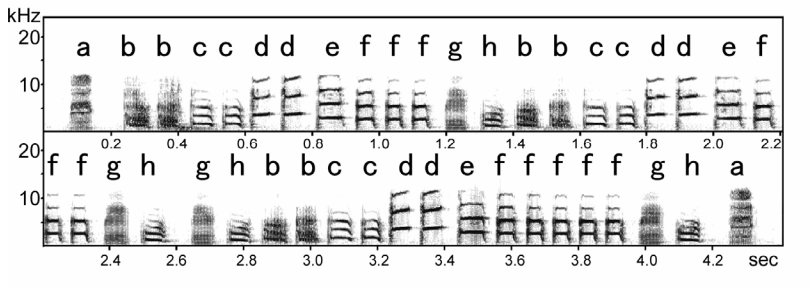

Fig.1 Sound spectrogram of the song note sequence of Bengalese finches.
Each note is identified as a discrete element on the sound spectrogram and is denoted by a letter of the alphabet.

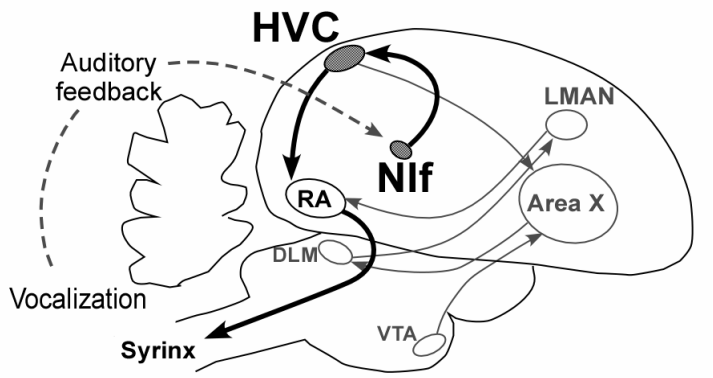

Fig.2 Neural basis of the birdsong referred to as the song system.
The NIf-HVC-RA pathway acts as the song production pathway. The other pathway (not highlighted) is responsible for song learning. LMAN: lateral magnocellular nucleus of anterior nidopallium, DLM: medial nucleus of the dorsolateral thalamus, VTA: ventral tegmental area.

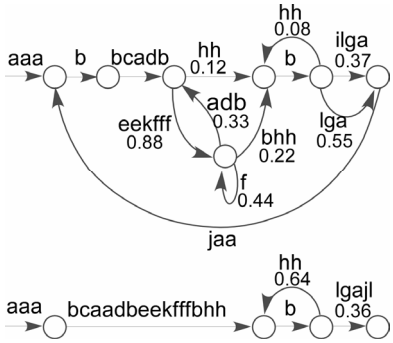

Fig.3 Changes in a song automaton as the result of a lesion in the NIf.
The upper and lower graphs correspond to pre- and post-lesion automata, respectively. Numerals indicate the probabilities of each branch. (Graphs are modified from Hosino and Okanoya (2000) with permission).

The nucleus interfacialis (NIf), one of the upstream parts of the HVC, is considered to be one of the essential regions that generate complexity in birdsongs. Lesions of the NIf reduce the branching of note-to-note transitions, however note sequences still correspond to paths on the original diagram (Fig. 3) (Hosino and Okanoya 2000). The reduction of complexity occurs only in birds having songs that are complex and not in birds that have simple songs. Based on this observation it is inferred that the NIf, in cooperation with the HVC, provides complexity for the generation of songs.

However, the types of interaction between the HVC and the NIf that can produce complex syntactical songs and the representation of the temporal patterns of songs in the HVC remain unclear. That these questions have not yet been answered is a consequence of the technical difficulties associated with investigating the actual interactions between brain regions of singing birds and developmental changes of such interactions.

In order to overcome these difficulties, a number of models have been proposed. For example, Fee's group proposed a model in which the HVC generates temporal patterns of songs by cooperating with the RA (Fee et al. 2004; Fiete et al. 2004). In this model, it is assumed that temporal patterns of songs are represented as feedforward activities of the HVC, the role of which is analogous to a recording "tape" (Fee et al. 2004). It is also assumed that these feedfoward activities innately exist and that the order of the notes in a song motif is acquired as a result of changes in connections between the HVC and the RA, where connections correspond to the mapping between an innately existing "tape" and each note sound. Doya and Sejnowski also developed a model based on the similar assumption (Doya and Sejnowski 1995). Although these models are sufficient for explaining the song generation of the Zebra finch, the songs of which are very linear, they are not sufficient to explain the song generation of the complex syntactical song of the Bengalese finch.

Only a few studies exist that consider the question of NIf function. Hoshino developed a statistical model of the song learning of the Bengalese finch using hierarchical Bayesian networks (Hoshino and Doya. 2000). In this model, the hierarchical structure of the note-chunk-motif is assumed to reflect the functional hierarchy of nuclei RA-HVC-NIf and the NIf is assumed to represent and control chunk-to-chunk transitions. However, the problem of this kind of model is that it is necessary to arbitrarily set a number of hyperparameters, for example different time constants for each hierarchy. Due to the difficulty of setting these parameters, the model was only successful in reproducing simple artificial model songs, and not in reproducing song sequences of the actual Bengalese finch.

The objective of the current study is to investigate the following two questions. First, what types of neural connectivity are able to generate the complex syntactical song sequences of the Bengalese finch, which is considered to potentially generate an infinite number of motif patterns? Secondly, what is the contribution of the NIf in song production of the Bengalese finch?

In order to describe complex temporal sequences like those of the Bengalese finch's song, which have hidden states and recursive structure, systems with temporal delay and recurrent connections are superior to the linear system proposed in the previous study. Hidden Markov models (HMM), which are equivalent to probabilistic finite state automata, are one of the most popular examples of this kind of system. However, because HMMs use symbol level abstraction, they are not sufficient for describing the neural basis for the observed phenomenon. On the other hand, although physiologically detailed models such as those that use spiking neurons have become popular recently, it is still difficult to reconstruct complex sequential behavior like that of the real Bengalese finch's song starting at the level of such spiking neurons.

In the current study, in order to mediate between a symbol-level computational model and physiologically detailed neural model, we propose a macro-level neural dynamics model for reproducing song production of the Bengalese finch. Focus of the current model is on how behavior of the Bengalese finch can arise from dynamics of neural connections representing groups of neurons in discrete brain nuclei.

With some modifications, a recurrent neural network (RNN) model, a type of model with feedback connections and time delays, is used as the main component of the current model. Because of its capacity to preserve internal state associated with complex dynamics, the RNN is often used for modeling of temporal sequence learning (Elman, J 1990, Jordan, MI and Rumelhart, DE 1992, Fetz EE and Shupe LE 2002). In a rate coding type neural network model, the model used in the current study, each unit's activity represents neural ensembles of groups of neurons, with dynamics, based on neural connectivity, providing spatiotemporal patterns of behavior. The RNN is thus considered to emulate characteristic features of actual neural systems, and the current model is considered consistent at the level of the macro-level mechanisms of biological neural systems.

For this reason, consistency in physiological details, such as features of neural activity at the level of individual neurons and characteristics of individual synapses, are not considered in detail. However, our macro-level model could easily be extended to a biologically precise model by adding different levels of modeling, such as for example a physiological model of individual neurons and synapses.

The model network is trained to generate song sequences of the Bengalese finch using template song sequences obtained from real birds. Results of simulations are analyzed through comparison with the song development of real young Bengalese finches, birds which learned the same song templates from their fathers.

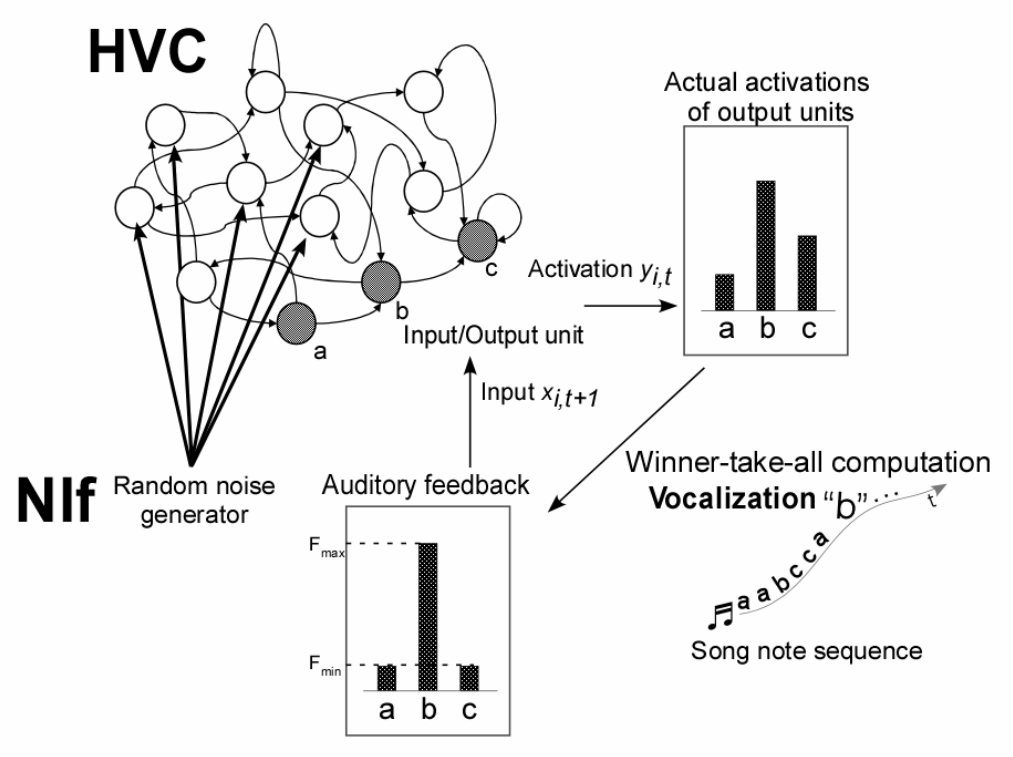

Fig.4 Model overview

Shaded circles are the output units of the RNN. Each output unit corresponds to a song note. Bars on the right top of the diagram indicate the activation of the output units at time *t*. Note "b" with the highest activation is selected to be the vocalized note sound at time *t*. The bars on the bottom of the diagram indicate the values of the output units after transformation, as dictated by the WTA computation. The value of the unit with the highest activation is set to F_{max}, all other output values of RNN are set to F_{min}. In the current experiment, for the normal feedback condition F_{max} is set to 0.8 and F_{min} is set to 0.2. The transformed values of output units act as the input at time $t+1$. The level of noise added to the non-output units is assumed to correspond to the activity of the NIf.

2. Model

2-1. Model overview

The architecture of the proposed model is shown in Figure 4. The HVC is modeled as a fully connected RNN that learns to generate temporal patterns of song note sequences. Every unit of the RNN is connected to every other unit, including itself. This connectivity allows the RNN to store internal states and to generate contextual sequences. Values of connection weights are asymmetric, i.e. the weight value from the *j*th unit to the *i*th unit (w_{ij}) is in general different from the weight value from the *i*th unit to the *j*th unit (w_{ji}).

The number of RNN units, including output units, is set to 35 for all learning trials. This is the minimum value large enough to successfully allow the network to learn songs with maximum number of notes. The number of output units is the same as the number of notes in the song that the model learns. Each of the output units corresponds to a note sound. Although outputs take on real number values, the output note of the model at each time step is selected by winner-take-all (WTA) computation: the output unit with the highest activation is selected as the output note of the network. The selection of the output note by the WTA computation corresponds to the vocalization process of songbirds, in the sense that real birds only generate whole notes like "a" or "b", not mixed sounds like "a+b". In the process of vocalization, WTA-like dynamics is considered to occur at downstream parts of the HVC, such as the RA.

In real birds, vocalized note sounds are sent back to the HVC via an auditory feedback process. To implement this process, the values of output units are replaced based on the vocalization process (the WTA computation). This transformation process, referred to as auditory feedback, is only applied to the output units in the generation mode. The set of outputs at time *t*, which is transformed according to the auditory feedback process, serves as the inputs for time *t+1*. The discrete time step of the RNN is incremented with each note output of song sequences.

The NIf is assumed to generate random noises, which are fed into the RNN units of the HVC. The noise provides stochasticity for branching of note sequences, in cooperation with the RNN of the HVC. The noise follows a uniform distribution and is added to the internal value of all units other than output units. The level of noise is defined in terms of the interval of the noise distribution. For example, if the noise level is set to 0.5 then the noise follows a uniform distribution on the interval [-0.5, 0.5]. Noise is added only during the generation of sequences, not during the model learning.

2-2. Generation mode

The model of neurons is a conventional firing rate model, in which the output of each unit is determined by applying a sigmoid function to the sum of all its inputs. The internal state ($u_{i,t}$) and the activation ($y_{i,t}$) of the ith unit at time t are determined by following formula

$$u_{i,t} = \begin{cases} \sum_{j \in N} w_{ij} x_{j,t-1} & i \in O \\ \sum_{j \in N} w_{ij} x_{j,t-1} + G & i \notin O \end{cases} \qquad (1)$$

$$y_{i,t} = f\left(u_{i,t} + \theta_i\right) \qquad (2)$$

where O is the set of indices corresponding to the output units, $x_{i,t}$ is the neural state of the ith unit at time t, N is the total number of units, θ_i is a threshold of the ith unit, $f(\,)$ is a sigmoidal function $f(x)=1/1+e^{-x}$, and G is the noise added to the internal value of non-output units. In the current sutdy, the value of G is set to 0.25 for all learning trials using six different template songs. This value is selected based on the average of optimal noise levels for all learning trials (c.f. Result).

After each time step, activation values of the non-output units, $y_{i,t}$, are simply copied to the neural states of next time step, $x_{i,t+1}$. Activation values of the output units, on the other hand, are transformed utilizing the WTA computation, described as follows.

$$x_{i,t+1} = \begin{cases} y_{i,t} & if \quad i \notin O \\ F_{\max} & if \quad i \in O \quad \wedge \quad y_{i,t} = \max_{j \in O} y_{j,t} \\ F_{\min} & otherwise \end{cases} \qquad (3)$$

Specifically, in the current study, the unit with the highest activation (F_{max}) is set to 0.8, whereas all other output values of the RNN (F_{min}) are set to 0.2. This limitation of the activation range of output units is basically performed in order to avoid divergence of weight values resulting from the sigmoid used in the activation function. This modification does not introduce any qualitative changes in results. The output of the network at time t is determined by selecting the note sound corresponding to the output unit with the highest activation level. The set of outputs at time t, transformed according to the WTA computation, serves as input for time $t+1$. The discrete time step of the RNN is incremented with each note output.

2-3. Training mode

In song learning of real birds, template song sequences are taken to be stored somewhere in the brain; the bird modifies its vocal output until the auditory feedback it receives matches the memorized template (Funabiki and Konishi 2003). Therefore in the proposed model, a network is trained by means of supervised learning using template song sequences obtained from real birds. The conventional back-propagation through time (BPTT) algorithm is used for learning of the model network (Rumelhart and McClelland 1986).

For the calculation of network activation in the training mode, additive noise G in equation (1) is set to 0. As in the case of the generation mode, after each time step, activation levels of the non-output units, $y_{i,t}$, are copied to the neural states of next time step, $x_{i,t+1}$. On the other hand, instead of the WTA computation of the output units, the desired activation value of output units at time t, $y^*_{i,t}$, serves as input for the next time step $t+1$ (open loop learning).

$$x_{i,t+1} = \begin{cases} y^*_{i,t} & i \in O \\ y_{i,t} & i \notin O \end{cases} \qquad (3')$$

The objective of learning is to find optimal values of connective weights that minimize the value of E defined as the learning error between the template sequences and output sequences. The error function E is described as follows,

$$E = \frac{1}{2T|O|} \sum_{i \in O} \sum_{t=1}^{T} \left(e_{i,t}\right)^2 \qquad (4)$$

$$e_{i,t} = \begin{cases} y_{i,t} - y^*_{i,t} & i \in O \\ 0 & i \notin O \end{cases} \qquad (5)$$

where T is the total length of teaching note sequences and $|O|$ is number of output units. According to the individual difference in songs of real birds, each teaching sequence has different numbers of output units and length. To normalize this difference, E is divided by T and $|O|$.

Connective weights can reach their optimal levels by updating their values in the opposite direction of the gradient $\partial E/\partial w$.

$$w_{ij}(n+1) = w_{ij}(n) - \alpha \frac{\partial E}{\partial w_{ij}} \qquad (6)$$

where α is the constant of the learning rate, and n is an index representing the iteration step in the learning process. $\partial E/\partial w$ is given by

$$\frac{\partial E}{\partial w_{ij}} = \sum_{t=1}^{T} f'\left(u_{i,t} + \theta_i\right) x_{j,t-1} \frac{\partial E}{\partial y_{i,t}} \qquad (7)$$

and is recursively calculated from the following recurrence formula

$$\frac{\partial E}{\partial y_{i,t}} = \begin{cases} \dfrac{1}{T|O|} e_{i,t} & i \in O \\ \displaystyle\sum_{k \in N} \frac{\partial E}{\partial y_{k,t+1}} f'\left(u_{k,t} + \theta_k\right) w_{ki} & i \notin O \end{cases} \qquad (8)$$

where $f'(\,)$ is the derivative of the sigmoid function. Through the iterative calculation of the BPTT, the values of connective weights reach their optimal values in the sense that the error between the template sequences and the output sequences is minimized.

Throughout the learning trials, the learning rate α is fixed at 1.0. The initial values of connective weights are set randomly to values ranging between -0.25 and 0.25. For both training and generation, initial states of the network are set to their neutral value, i.e. the internal state of each neuron is set to 0.

Template	Variety of notes	Total No. of notes	Block entropy	Example of note sequence
Song A	8	604	1.44	bccddeefaaaghghbccddeefaaagh
Song B	10	604	1.55	aibbbcdefhccdefajggghccaibbcd
Song C	10	695	1.41	abbccbadeefghabijjabbccbadeef
Song D	12	660	1.74	abcdefabcdefggiabchfdefhfijcm
Song E	8	781	1.42	hhafagabacdedafagfhhagabacde
Song F	6	619	1.02	abccdefbccdefbccdefbccdefbcc

Table.1 Template songs used for the six learning trials.

The songs are sampled from six different family lines and are described by sequences of letters. Each song differs in the number of note elements and in the diversity of note-to-note transition rules. Approximately ten song bouts for each template song are used for the learning of the model. Differences in the total number of notes is caused by differences in the length of each song bout.

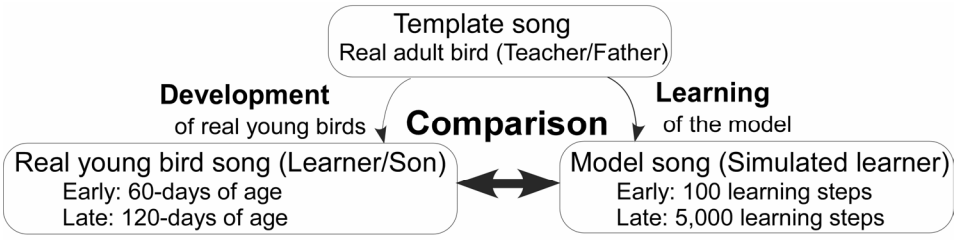

Fig. 5. Experimental procedure.

Song sequences of a real adult bird (teacher/father) are used as the template for the learning of the model. Song sequences of the model (simulated learner) in the learning process are compared with song sequences of a real young bird (learner/son) that learned from the same template song.

3. Experiments

3-1. Experimental procedure

Six learning trials are attempted using six different songs of real adult birds as teacher signals. The songs of the animal subjects are different from each other because the birds come from different family lines (Table.1). The songs of each animal subject were recorded for one hour, and all of them were used for training of the model as teacher signals. The number of song bouts in each recorded song varies from 8 to 10, wiht each consisting of about 60 to 80 notes. Thus the total number of notes in each subject bird's songs varies from 604 to 781. Songs of real birds are converted to sequences of letters through sound spectrogram analysis.

To compare the development of real birds with the learning process of the model, song sequences in the development of real young Bengalese finches that learned the same templates are also sampled. For each young bird, song note sequences are sampled at approximately 60-days of age when song notes are first identifiable and again at approximately 120-days of age when songs are crystallized.

The sequences generated by the model are also sampled twice in the early (100 steps) and late (5,000 steps) stages of the learning process. Both in the early and late stages of learning, weight values are fixed and 100 sequences of 50 notes each are generated with noise added. The amount of noise is varied to investigate the role of noise.

To evaluate the learning processes and the performance of the model, the sequences generated are compared with the songs of real young Bengalese finches that learned the same template song (Fig. 5). Song note sequences of real young birds and those of the model are evaluated using measures determined as described in the following section.

3-2. Analysis measure

Analyses are conducted by counting the occurrence frequencies of letter blocks in each song sequence. The distribution of letter block occurrence probabilities reflects distinguishing features of each song. According to the previous study, the songs of the Bengalese finch can usually be satisfactorily reproduced using a third-order Markov model (Hosino and Okanoya 2000). Based on this, we examined letter blocks of length between 3 and 7. Results, however, did not show any qualitative difference. Also because the amount of actual birdsong data is limited, the longer the length of letter block is, the less reliable the calculated statistics are. Therefore, in the present study, results for a letter block length of 3 are shown.

In order to evaluate the similarity between two song sequences, the Kullback-Leibler divergence (KL-divergence), a well-known distance measure of probabilistic distributions, is used (Cover and Thomas 1991). The KL-divergence is determined by the following formula:

$$d_{KL}(D,D') = \sum_{x \in A^n} P_D(x) \log \frac{P_D(x)}{P_{D'}(x)} \qquad (9)$$

where A corresponds to the component notes of a particular song, A^n is the set of all strings of length n that can be built from A, D and D' correspond to a particular song syntax, which has a probabilistic distribution of strings, and $P_D(x)$ is the occurrence probability of string x under distribution D. Specifically, in the current study, D and D' correspond to syntaxes of the template song and the learner's song, respectively. As is standard, we set 0log0=0 and 0/0=1. In cases where there is a string with a null probability in D', but not in D, $P_{D'}(x)$ is set to a small value (1.0×10^{-6}) to avoid division by zero.

To evaluate the diversity of transitions in songs, we use the block entropy, determined by the following formula

$$H_n = -\frac{1}{n} \sum_{x \in A^n} P_D(x) \log P_D(x) \qquad (10)$$

In both cases, n is set to 3.

Letter blocks not appearing in the template are ignored in the in the calculation of KL-divergence. Therefore to measure how many seemingly random transitions appear in the learner's sequences, we also calculate the occurrence rate of these letter blocks (out of template).

In order to investigate the influence of the occurrence rate of letter blocks in the template on the song produced by the learners, we calculate how many, of the ten most frequently appearing letter blocks in the template, are reproduced in the songs of the learners (10 most frequent blocks).

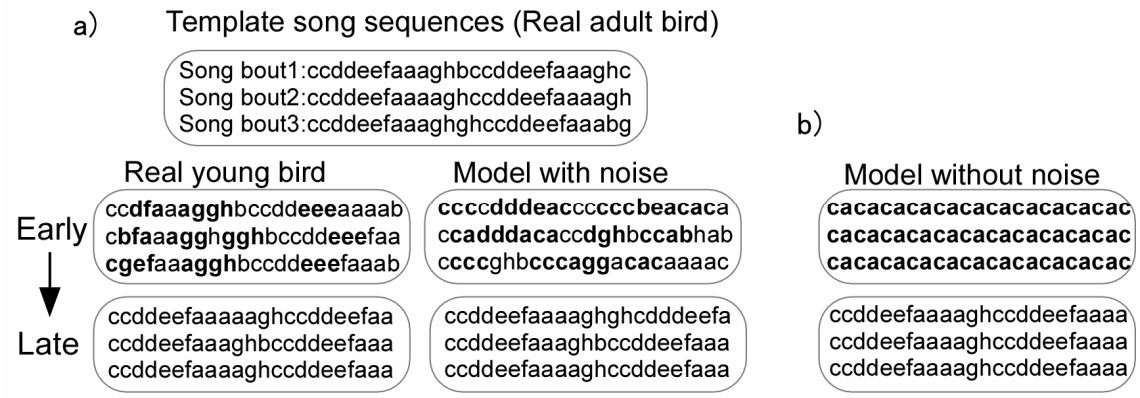

Fig.6 Example of a learned template, song note sequences of a real young bird in development, and song note sequences generated by the model (a) with noise and (b) without noise during the learning process.

The top row illustrates a template song (Song A) learned by a real young bird and the model. The early stage (middle row) illustrates song note sequences of a 61-day-old bird in the case of a real young bird, and 100 steps of learning in the case of the model. The late stage (bottom row) illustrates song note sequences of a 127-day-old bird in a real bird and 5,000 steps of learning in the case of the model. In the simulations with noise, the nose level is set to 0.25. Bold letters indicate letter blocks that do not appear in the template.

4. Results

4-1. Performance of the model

Performance of the model is evaluated by calculating the KL-divergence between the template songs and the learners' songs. The model is trained to imitate the song sequences of the Bengalese finch using songs of real adult birds as teacher signals. As a result of the learning process, KL-divergence decreases until it reaches a level that corresponds to the individual deviations typical of young Bengalese finches (Fig. 7a). This demonstrates that the proposed model successfully learns to generate complex syntactical songs nearly identical to those of young Bengalese finches.

In the case of the real birds, the early stage of development is defined as the moment when song notes are first identifiable. Even at this early stage, there are some birds whose KL-divergence has decreased to some extent. This indicates that in the case of real birds the learning of notes and learning of note-to-note transition rules occur in parallel with each other. Determination of the early stage in the learning process of the model is difficult, because the process of note learning is not implemented. In the current study, the early stage is determined as 100 steps of learning for all learning

trials. This value is selected since KL-divergence between the model and the templates is most similar to the KL-divergence between the real young birds and their fathers at this early stage.

4-2. Comparison of the development of a real bird and the learning process of the model

Figure. 6a. shows an example of a learned template, song note sequences of a real young bird in development, and song note sequences generated by the model with noise over the learning process. It may be observed that the template song consists of probabilistic combinations of the chunks "ccddee", "faaa", "gh" and the note "b". When they appear in direct succession, the chunks "ccddee" and "faaa" (which together form the chunk "ccddeefaaa") are often followed by "gh", but sometimes branches to "b". After the note "b", song sequences again branch to "ccddee" and "gh". This is a temporal song structure typical of the Bengalese finch.

At the beginning of the learning process, both in the songs of real birds and in the songs of the model, many transitions appear that are not present in the template song (bold letters). However, letter blocks that frequently appear in the template, for example "ccd", "cdd", "faa" and "aaa", which construct chunks "ccddee" and "faaa" of the template, are already reproduced at this early stage.

By the end of the learning process, both the model and the real young bird can almost completely replicate the features of the template song. Letter blocks that do not appear in the template disappear. Moreover, letter blocks that correspond to the branching points of chunks, such as "bcc","bgh" and "ghg", begin to be reproduced with probabilistic distributions similar to those of the template sequences. This observation suggests that the real birds first learn and reproduce chunks that most frequently appear in the template, and then next begin to reproduce chunk-to-chunk combinations with probabilistic distributions. These trends are also observed in the learning process of the model.

The similarity of trends between the development of the real bird and the learning of the model is also observed in the study of statistical measures. In the early stages of both the development of the real bird and the learning process of the model, there are many letter blocks that are seemingly random and do not appear in the template sequences (Fig. 7(c)). Thus the variety of generated sequences is initially large (Fig. 7(b)). However, at the same time, letter blocks that appear frequently in the template sequences are reproduced at this very early stage (Fig. 7(d)).

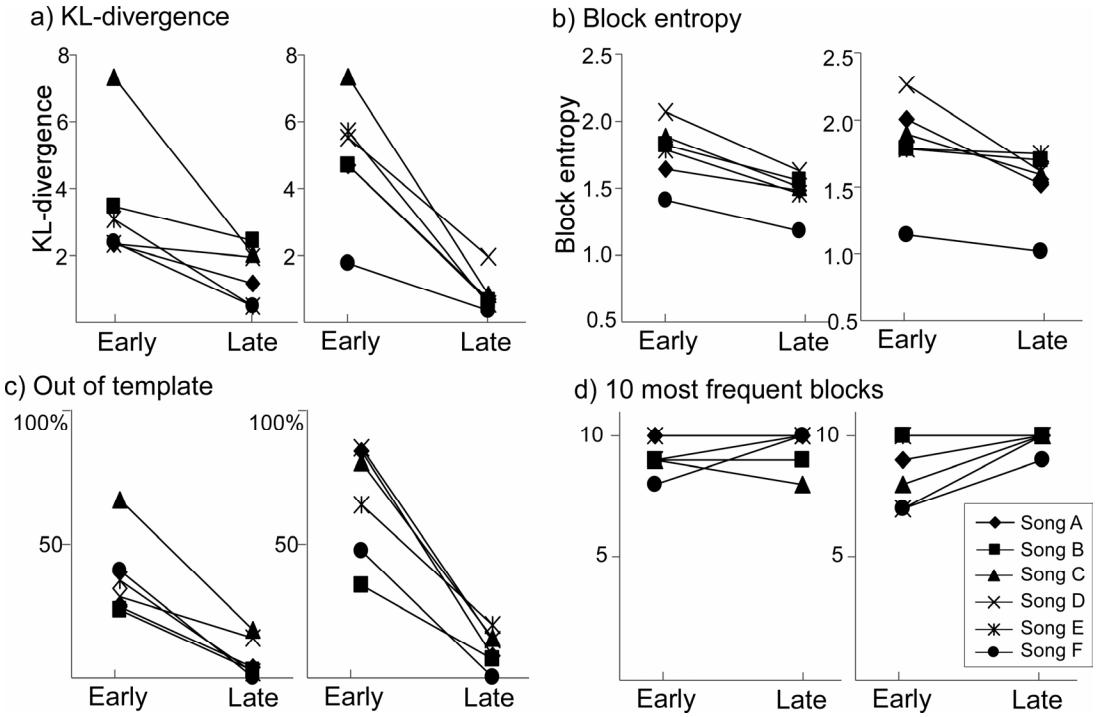

Fig.7 Comparison of the development of real birds and the learning process of the model.

For each measure studied, the left-hand graph corresponds to real birds, whereas the right-hand graph corresponds to the model. (a) KL-divergence values at early and late stages of each process. (b) Changes in the block entropy. (c) Changes in the occurrence rates of the letter blocks that do not appear in the template sequences. (d) Changes in the 10 most frequent letter blocks of the template sequences. The amount of noise is fixed at 0.25 for all learning trials both in the early and later stages of learning.

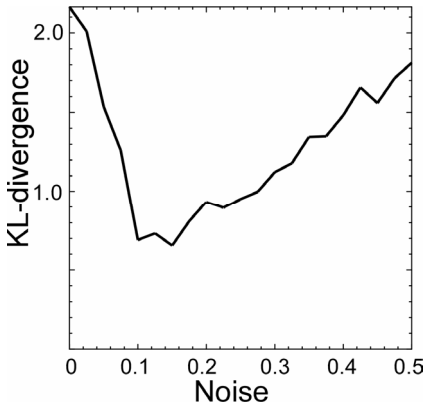

Fig.8 Example of the relationship between the KL-divergence and the level of noise added (Song B).

The graph indicates that there is an optimal noise range within which the model generates note sequences that have branching probability distributions very similar to the template.

By the end of development and learning, letter blocks whose occurrence probabilities are relatively low are eventually also reproduced. Seemingly random transitions and letter blocks that do not appear in the template almost completely disappear (Figs. 7(c)). As a result, the diversity of the generated sequences is lower than the diversity at the early stages of each process (Figs. 7(b)). In terms of these measures, changes in the development of the real birds and in learning process of the model show similar trends. That is, the real birds first learn and reproduce chunks, and then begin to reproduce chunk-to-chunk combinations with probabilistic distributions. This suggests that the developmental learning process of note-to-note transition rules in real birds is similar to the learning process of the model, where the activation of the network represents probabilistic distributions of template sequences as dynamical systems.

In the case of the model, these features of the learning process can be explained in terms of changes in the effects of noise resulting from the changing structure of the network. In the learning process, the network dynamics of the model reproduce note-to-note transitions in probability form by updating connective weights to minimize the error between generated sequences and the teacher signal.

At the beginning of the learning process, since the connective weights have not yet converged, the effect of noise is large, and generated sequences seem random. In the process of learning, letter blocks that frequently appear in the templates are reproduced first. However, seemingly random transitions are still observed. As the structure of the network converges, the effect of noise gradually diminishes, and blocks that do not appear in the template eventually disappear. By the end of the learning process, the activation of the RNN represents the occurrence probabilities of each branch, as given by the template sequences. Therefore, the role of noise is to provide stochasticity for branching. Even less frequent letter blocks in the template are eventually replicated with occurrence probabilities close to that of the template.

4-3. Function of noise

To examine the role of noise in the model, the connective weights of the networks are fixed at the end stage of learning, and sequences are generated with various amounts of noise added. Figure 8 illustrates the relationship between the KL-divergence and the level of noise added for one trial (Song B). This curve shows that the KL-divergence is large both for extremely small and extremely large amounts of noise and that it reaches a minimum for intermediate amount of noise. This indicates that there is an optimal noise range within which the model generates note sequences that have branching probability distributions very similar to the template.

Comparison of block entropies for sequences with and without noise reveals that, for songs with higher entropies, the entropy is reduced by the removal of noise. In contrast, for the low-entropy song (Song F), there is almost no effect (Fig. 9). This result is consistent with the previous NIf lesion study in that the entropy of high-entropy songs is reduced, whereas, in the case of low-entropy songs, almost no effect is observed.

The results of a simulated NIf lesion study can also be explained in terms of changes in the effects of noise on the network. At the end of the learning process, once weights are fixed, there is an optimal noise range within which the model generates note sequences that have branching probability distributions that are very similar to the template (Fig. 8). If the amount of noise is increased beyond the optimal range, seemingly random transitions appear, which do not occur in the template sequence. If the amount of noise is decreased from the optimal range, branches with lower occurrence probabilities disappear, and eventually only the most frequent path remains, despite the fact that the note sequences learned from the template contain some branching. However, in the case of a song with almost no branches, decreasing the amount of noise has almost no effect. Thus, the network that learns from a simple song template is not affected by the removal of noise.

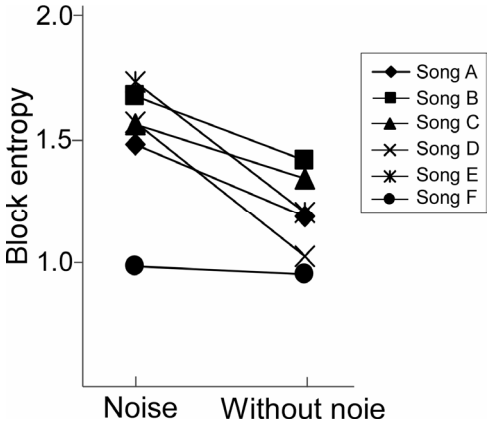

Fig.9 Comparison of block entropies on sequences with and without noise.

In the simulations with noise, the nose level is set to 0.25. For the songs with higher entropies, the entropy is reduced by the removal of noise. In contrast, for the low-entropy song (Song F), there is almost no effect.

5. Discussion

5-1. Representation of the temporal pattern of song in the HVC

Our biological observations on the song development process of real birds show that at the beginning of the learning process, real birds first learn and reproduce chunks that most frequently appear in the template, and then next begin to reproduce chunk-to-chunk combinations with probabilistic distributions. Even less frequent letter blocks in the template are eventually replicated with occurrence probabilities close to that of the template. This biological observation suggests that the previous model of temporal pattern acquisition in the HVC might not be applicable in the case of the Bengalese finch.

In the previous model developed by Fee's group, it is assumed that temporal patterns of songs are represented as feedforward activities of the HVC, the role of which is analogous to a recording "tape" (Fee et al. 2004). It is also assumed that these feedfoward activities innately exist and that the order of the notes in a song motif is acquired only as a result of changes in connections between the HVC and the RA, where connections correspond to the mapping between an innately existing "tape" and each note sound. They also hypothesized that birds whose song repertoire size is large have a large number of tapes (Fee et al. 2004).

According to the assumptions made in their model, a bird which generates an almost infinite number of different motifs, such as the Bengalese finch, should have an almost infinite number of tapes encoding these motifs. However, if certain chunks are independently generated based on different neural activities, even in cases where the final output sounds correspond to the same particular chunk, in Fee's model it is hard to explain the "chunking" phenomenon which is observed.

In the current proposed model, however, particular chunks correspond to certain neural states in a generalized manner, in the sense that particular chunks are encoded in one particular neural state even when they appear in different motifs. The appearance of particular chunks in the learning process is interpreted in a consistent way as resulting from certain neural states corresponding to certain chunks developed in the HVC. Therefore, biological observations suggest that the temporal note sequences of songs would take the form of recurrent dynamics of the kind that have been shown in our model, rather than feedforward activities, the mechanism proposed in the previous model.

5-2. Function of the NIf

In the current model, additive noise, in cooperation with deterministic dynamics of the HVC, is enough to reproduce occurrence probabilities of branching in song note sequences, even though noise is provided from the NIf independent of the note sequence context. Moreover, the simulated functional change of the NIf also produces results that are consistent with a previous lesion study.

In the proposed model, activation of the HVC represents occurrence probabilities of each branch in song note sequences, with random activity of the NIf providing stochasticity for branching. Thus, reduction of noise reduces the level of branching and results in a decrease in the entropy of songs. In addition to changes in the entropy of songs, the model also predicts which particular path in the song note sequences should disappear as a result of the reduction of NIf activity. Specifically, if the amount of noise is decreased from the optimal range, branches with lower occurrence probabilities disappear first, and eventually only the most frequently traversed path remains.

Without noise, variability of songs is substantially limited, and the model repeatedly produces the same linear note sequence (Fig 6b). Lesions to the NIf, however, normally result in complete and irreversible inhibition of activity. Thus, in order to verify the prediction, instead of a lesion study we are conducting a neurophysiological experiment by using chemicals to reversibly control the activity of the NIf.

In the previous model, song diversity of the Bengalese finch is assumed to be provided by more complex, higher-level functions of the NIf corresponding to functional hierarchy of the nuclei RA-HVC-NIf (Hoshino and Doya. 2000). However the proposed model shows that only context independent noise is sufficient to reproduce occurrence probabilities of branching in song note sequences. Therefore, the present model suggests that the NIf may not need such a complex representation.

Our hypothesis is consistent with the fact that the NIf is a very small nucleus consisting of a small number of neurons and the fact that the activation pattern of the NIf is less context-dependent than the HVC and the RA (McCasland 1987). This is also consistent with the fact that the Bengalese finch and the Zebra finch, the songs of which are simple, are closely related and may indicate that there are no major functional differences between these species.

Moreover, the assumption that the NIf performs a noise-like function suggests a possible connection with the perspective adapted in the study of reinforcement learning. We hypothesize that the noise-like activity of the NIf could help exploration in the learning of note-to-note transition rules by generating fluctuations at each transition. This is similar to the hypothesis that states that the random activity of the lateral

magnocellular nucleus of anterior nidopallium (LMAN) helps exploration in note learning by fluctuating each individual sound (Ölveczky et al. 2005). Although the current model is trained using supervised teaching, in the near future, we plan to use reinforcement learning to further investigate the hypothesis regarding the possible role of noise-like activity of the NIf.

As shown earlier, each template song has an optimal noise range, within which the model generates note sequences with branching probability distributions that are very similar to the template. In the present study, although the amount of noise is fixed for all individual birds and for each learning period, we are able to enable the model to learn this optimal range by allowing the amount of noise to vary, in parallel with changes in connective weights. If the song generation of real birds is dependent on such a noise range, then we can speculate that NIf activity changes during the developmental process and that these dynamics may be reflected in differences between the Bengalese finch and the Zebra finch. That is, at the early stage of learning note-to-note transition rules, higher activity of the NIf is needed for exploration, whereas this activity is not necessary at the end of the learning process. In the Bengalese finch, however, NIf activity is relatively high at the end of the learning process, allowing the Bengalese finch to learn to sing complex songs with a syntactical structure. To confirm these predictions, changes of NIf activity during development and differences in NIf activity between the Bengalese finch and the Zebra finch need to be examined.

6. Conclusion

We proposed a novel hypothesis based on biologically supported assumptions that provides an explanation for the functional role of NIf-HVC interaction that generate complex syntactical songs of the Bengalese finch. The model shows that complex syntactical songs, described by a probabilistic finite state syntax, can be replicated by simple interactions between deterministic dynamics of a recurrent neural network and random noise. Features of the learning process in the simulations show similar trends to empirical data on the song development of real birds. This observation suggests that the temporal note sequences of songs take the form of a dynamical process involving recurrent connections in the network of the HVC, as opposed to feedforward activities, the mechanism proposed in the previous model.

Moreover a functional change simulating a NIf lesion induced by manipulating interactioins between determinisitc dynamics and random noise produces results that are consistent with a previous empirical lesion study of the NIf. The model also provides a number of testable hypotheses that could contribute to further studies of NIf functions.

References

Cover,TM., Thomas, JA. (1991). Element of Information Theory. NewYork, Wiley.

Doya, K., and Sejnowski, T.J. (1995). A novel reinforcement model of birdsong vocalization learning. Advances in Neural Information Processing Systems, 7, 101-108.

Elman, J. (1990). Finding structure in time. Cognitive Science 14: 179-211.

Fee, MS., Kozhevnikov, AA., Hahnloser, RH. (2004). Neural mechanisms of vocal sequence generation in the songbird. Ann N Y Acad Sci, 1016, 153-70.

Fetz EE and Shupe LE. (2002). Recurrent network: Neurophysiological modeling. In: The hand book of brain theory and neural network. MIT Press, Cambridge.

Fiete, IR., Hahnloser, RH., Fee, MS., Seung, HS. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. J Neurophysiol, 92, 2274-82.

Honda, E., and Okanoya, K. (1999). Acoustical and syntactical comparison between songs of the white-backed Munia(Lonchura striata) and its Domesticated strain, the Bengalese finch(Lonchura striata var.domestica). Zool Sci, l16, 319-326.

Hosino, T. and Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. Neuroreport, 11, 2091-2095.

Hoshino, C. and Doya, K. (2000). Learnig of Hierarchical Temporal Sequences in Biological Systems. *Technical Report of IEICE,* NC99-108, 117-124.

Jordan, MI. and Rumelhart, DE. (1992) . Forward models: supervised learning with a distal teacher. Cognitive Science 16, 307-354.

McCasland, JS. (1987). Neuronal control of bird song production. J Neuroscience, 7, 23-39.

Nottebohm, F. (2005). The neural basis of birdsong. PLoS Biology, l3, 759-761.

Nottebohm, F., Stokes, TM. and Leonard, CM. (1976). Central control of song in the canary, Serin us canaries. The Journal of comparative neurology, 165, 457-86.

Ölveczky, BP., Andalman, AS. and Fee, MS. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biology, l3, 902-909.

Rumelhart DE. and McClelland JL. Parallel Distributed Processing. (1986). Parallel Distributed Processing. MIT Press, Cambridge

Vu, ET., Mazurek, ME., Kuo, YC. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. J Nneuroscience, 14, 6924-6934.

Zann, RA. (1996). The Zebra Finch. Oxford University Press, Oxford