Hidden Markov processes can explain complex sequencing rules of birdsong: A statistical analysis and neural network modeling

Kentaro Katahira\textsuperscript{1,2,3}, Kenta Suzuki\textsuperscript{3,4}, Kazuo Okanoya\textsuperscript{1,2,3}, and Masato Okada\textsuperscript{1,2,3}

1. JST ERATO, Okanoya Emotional Information Project, 2. The University of Tokyo, 3. RIKEN Brain Science Institute, 4. Saitama University
Motivation
- What are neural substrates for sequential behavior?

Sequential behavior
- Speech
- Playing music
- Dancing

Generation

Perception

Learning
Motivation
- What are neural substrates for sequential behavior?

Birdsong

Generation
Perception
Learning

Syllable: a b c d
Frequency
1. Introduction
   - Neural substrates of birdsong
   - Neural network models
2. Statistics of birdsong
   - Higher-order history dependency
3. Statistical models for birdsong
4. Discussion
   - Neural implementation
   - Future directions
Neural activity pattern during singing

Hahnloser, Kozhevnikov and Fee, Nature, 2002
Feedforward chain hypothesis

- Spikes propagate on feedforward chain network


Experimental evidences:

It is suitable for fixed sequences.
But how about variable sequences?
Song of Bengalese finch
- Variable sequences including branching points
Branching-chain hypothesis

- Mutual inhibition between branching chains

(Jin, Phys Rev E, 2009)
Limitation of branching-chain model

- The transition is a simple Markov process
  - The present active chain depends only on the last active chain

Question: Syllable sequences of Bengalese finch songs are Markov processes?
Outline

1. Introduction
   – Neural substrates of birdsong
   – Neural network models
2. Statistics of birdsong
   – Higher-order history dependency
3. Statistical models for birdsong
4. Discussion
   – Neural implementation
   – Future directions
Test of (first order) Markov assumption

Null hypothesis:
The transition probability to next syllable does not depend on preceding syllable (Markov assumption)

\[ \begin{align*}
\text{Prob.} & \quad \text{Prob.} \\
b & 0.385 \quad 0.385 \\
c & 0.495 \quad 0.495 \\
d & 0.408 \quad 0.422 \\
e & 0.097 \quad 0.193 \\
\end{align*} \]

\( \chi^2 \) goodness-of-fit test

Significant difference

Second-order history dependency
We found more than one significant second-order history dependency in all 16 birds. 
($p < 0.01$ with Bonferroni correction)
Then,…

• The branching-chain model is incorrect?
Two possible mechanisms for history dependency

Hypothesis 1:
Chain transition with higher-order dependency

Hypothesis 2:
Many-to-one mapping from chains to syllables

(Katahira, Okanoya and Okada, Biol. Cybern. 2007)
However...

- The neural activity data from HVC of *singing Bengalese finches* are *not* available.

- We examined two hypotheses based on song data by using statistical models.
Outline

1. Introduction
   – Neural substrates of birdsong
   – Neural network models
2. Statistics of birdsong
   – Higher-order history dependency
3. Statistical models for birdsong
4. Discussion
   – Neural implementation
   – Future directions
Feature extraction - Auditory features

Auditory features
- Spectral entropy
- Duration
- Mean frequency

(c.f. Tchernichovski et al. 2000)
Hidden Markov Model (HMM)

\[ p(x|y = i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right\} \]
State transition dynamics in HMM

1\textsuperscript{st} order HMM: \(a_{ij} = p(y_t = j | y_{t-1} = i)\)

2\textsuperscript{nd} order HMM: \(a_{ijk} = p(y_t = k | y_{t-1} = j, y_{t-2} = i)\)

0\textsuperscript{th} order HMM (Gaussian mixture):
\(a_i = p(y_t = i)\)
Relationship between two hypotheses and statistical models

Hypothesis 1:
Chain transition with higher-order dependency

Hypothesis 2:
Many-to-one mapping from chains to syllables

2nd order-HMM

1st order-HMM
Bayesian model selection

Given data (auditory features): \( X = \{x_1, x_2, \ldots\} \)

Model structure \( \mathcal{M} = \{L, K\} \)
- \( L \): Markov order (0, 1, 2)
- \( K \): the number of hidden states

Model posterior: \( p(\mathcal{M} | X) \propto p(X | \mathcal{M}) p(\mathcal{M}) \)

Marginal likelihood: \( p(X | \mathcal{M}) = \int d\theta p(X | \theta, \mathcal{M}) p(\theta | \mathcal{M}) \)
\( (\theta: \text{model parameter set}) \quad (\ difficul to compute!) \)

Approximation
\[
\log p(X | \mathcal{M}) \geq \mathcal{F}_\mathcal{M} \quad \text{Lower bound}
\]
\( (\text{variational free energy}) \)
\( (\text{can be computed by variational Bayes method}) \)
Result – model selection (one bird)

“Best model structure”

- With small number of states
  - 2nd order HMM
- With large number of states
  - 1st order HMM

Number of states, K

Lower bound on log-marginal likelihood

- With **small** number of states  □  2nd order HMM
- With **large** number of states  □  1st order HMM
Results – model selection, cross validation
(averages over 16 birds)

Lower bound on log-marginal likelihood

Predictive likelihood (cross validation)
HMM learns many-to-one mapping

(Similar results were obtained for 30 syllables of the 54 syllables where significant second-order dependency was found)
Outline

1. Introduction
   – Neural substrates of birdsong
   – Neural network models
2. Statistics of birdsong
   – Higher-order history dependency
3. Statistical models for birdsong
4. Discussion
   – Neural implementation
   – Future directions
Summary of results

- Bengalese finch songs have at least second-order history dependency.

This mechanism is sufficient for Bengalese finch song
Mapping onto neuroanatomy

- **HVC** - hidden state (branch state)
- **RA** - auditory features of each syllable

(Katahira, Okanoya and Okada, 2007)
Future directions (ongoing research)

- How the brain can learn this representation?
  - Analysis of development of song from a juvenile period.
  - Developing a network model with synaptic plasticity for learning the many-to-one mapping. (e.g., Doya & Sejnowski, NIPS, 1995; Troyer & Doupe, J Neuropysiol, 2000; Fiete, Fee & Seung, J Neuropysiol, 2007)

- Applying HMMs to spike data recorded from songbird (Katahira, Nishikawa, Okanoya & Okada, Neural Comput, 2010)
Overview of our approach

Neural network model

Constraints

Behavior

Parameter fitting, Model selection

Constraints

Support, Refinement

Statistical model

Mapping

Anatomy, Physiology