A method of inferring partially observable Markov models from syllable sequences reveals the effects of deafening on Bengalese finch song syntax

Jiali Lu[1†], Sumithra Surendralal[2†], Kristofer E. Bouchard[3,4], and Dezhe Z. Jin[1*]

1 Department of Physics and Center for Neural Engineering, The Pennsylvania State University, University Park, PA, USA
2 Symbiosis School for Liberal Arts, Symbiosis International (Deemed University), Pune, Maharashtra, India
3 Scientific Data Division and Biological Systems & Engineering Division, Lawerence Berkeley National Laboratory
4 Helen Wills Neuroscience Institute & Redwood Center for Theoretical Neuroscience, UC Berkeley

† These authors contributed equally to the project
∗ Corresponding author: Dezhe Z. Jin, dzj2@psu.edu

Abbreviated title: **POMMs for Bengalese finch songs**

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: Research was supported by NSF award EF-1822476 (DZJ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Abstract

Songs of the Bengalese finch consist of variable sequences of syllables. The sequences follow probabilistic rules, and can be statistically described by partially observable Markov models (POMMs), which consist of states and probabilistic transitions between them. Each state is associated with a syllable, and one syllable can be associated with multiple states. This multiplicity of syllable to states association distinguishes a POMM from a simple Markov model, in which one syllable is associated with one state. The multiplicity indicates that syllable transitions are context-dependent. Here we present a novel method of inferring a POMM with minimal number of states from a finite number of observed sequences. We apply the method to infer POMMs for songs of six adult male Bengalese finches before and shortly after deafening. Before deafening, the models all require multiple states, but with varying degrees of state multiplicity for individual birds. Deafening reduces the state multiplicity for all birds. For three birds, the models become Markovian, while for the other three, the multiplicity persists for some syllables. These observations indicate that auditory feedback contributes to, but is not the only source of, the context dependencies of syllable transitions in Bengalese finch song.

Author Summary

Context dependencies are widely observed in animal behaviors. We devise a novel statistical method for uncovering context dependencies in behavioral sequences. Application of the method to songs of the Bengalese finch before and shortly after deafening reveals that auditory feedback contributes significantly to context dependencies, but is not the only source. Our approach can be applied to many other behavioral sequences and aid the discovery of the underlying neural mechanisms for context dependencies.

Introduction

Consisting of sequences of stereotypical syllables, birdsong has numerous parallels with human speech (Doupe and Kuhl, 1999). Syllable sequences of many songbird species are variable, and follow probabilistic rules (or syntax) that can be described with state transition models (Okanoya, 2004; Jin and Kozhevnikov, 2011; Jin, 2013; Markowitz et al., 2013). For Bengalese finch songs, it was shown that the syllable sequences are well described by partially observable Markov models (POMMs) (Jin and Kozhevnikov, 2011). In a POMM, the state transitions are Markovian: the transition probabilities between the states are fixed and do not depend on the history of the state transitions. Each state is associated with one syllable. This enables a POMM to generate syllable sequences through the state transitions. Although a state is associated with one syllable, the converse is not necessarily true. In a POMM, one syllable can be associated with multiple states. This multiplicity of syllable to states association enables a POMM to describe context dependences in syllable transitions: transition probabilities between syllables depends on the preceding syllable sequences (Jin and Kozhevnikov, 2011). The Markov model is a special case of POMM, in which there is one-to-one correspondence between the states and the syllables. Markov models are not capable of describing context dependencies in syllable transitions.

POMM is motivated by the idea that birdsong is driven by synaptic chains in the premotor nucleus HVC (proper name) of the song system (Hahnloser et al., 2002; Fee et al., 2004; Jun and Jin, 2007; Jin et al., 2007; Jin, 2009; Long et al., 2010; Wittenbach et al., 2015; Lynch et al., 2016; Picardo et al., 2016; Jin, 2013; Zhang et al., 2017; Egger et al., 2020; Tupikov and Jin, 2021). Specifically, the HVC neurons that project to the downstream motor areas form feedforward synaptic chain networks within HVC. Bursts of spikes propagate along a chain, with each projection neuron bursting once during the propagation, driving the production of one syllable through the projections to the downstream motor areas (Fee et al., 2004; Jin, 2009). The activation of one such “syllable-chain” can be identified as the neural correlate of one state in a POMM (Jin, 2009; Jin and Kozhevnikov, 2011; Wittenbach et al., 2015). Within this paradigm, inferring POMMs from observed syllable sequences can shed light on the neural dynamics in HVC that underlies production of variable syllable sequences.

Auditory feedback has been shown to affect Bengalese finch song syntax (Okanoya and Yamaguchi, 1997; Woolley and Rubel, 1997; Woolley and Rubel, 2002; Sakata and Brainard, 2008; Wittenbach et al., 2015). A few days after deafening, the syllable sequences become more random (Okanoya and Yamaguchi, 1997; Woolley and Rubel, 1997), and the number of repetitions of long repeating syllables become smaller (Wittenbach et al., 2015). Altered auditory feedback to intact singing birds delivered at branching points of syllable transitions can change the transition probabilities (Sakata and Brainard, 2006; Sakata and Brainard, 2008). These observations demonstrate that auditory feedback could play an important role in creating context dependencies in syllable transitions in Bengalese finch song.

In this paper, we analyze the songs of six Bengalese finches before and shortly after deafening. We first devise a novel method for inferring a POMM from a set of observed syllable sequences. The method depends on the concept of *sequence completeness*, which is the total probability that the POMM generates all of the unique sequences in the observed set. Sequence completeness is further augmented with the differences of the probabilities of the unique sequences computed with the observed set or with the model, leading to the augmented sequence completeness, P_β. The method is designed to find the minimum number of states for each syllable such that P_β of the observed sequences is statistically compatible with the POMM. Compared to the previous heuristic method of inferring POMMs from observed syllable sequences (Jin and Kozhevnikov, 2011), our new method is much simpler and more principled.

Using this method, we infer minimal POMMs for the syllable sequences of the birds before and after deafening. We show that deafening reduces the number of states required in the POMMs, indicating that deafening reduces context dependencies in the syllable transitions. Before deafening, the POMMs of all birds require multiple states for some syllables. After deafening, the POMMs are reduced to simple Markov models for three birds, while for the remaining three the multiplicity of the states persists for some syllables. Our results indicate that auditory feedback contributes to context-dependent syllable transitions, but other mechanisms such as multiple syllable-chains encoding the same syllable should also contribute (Jin, 2009; Jin and Kozhevnikov, 2011; Cohen et al., 2020).

Results

In this paper, we analyze the dataset collected for a previous study (Wittenbach et al., 2015), which showed that syllable repeats in Bengalese finch songs, especially for those syllable types with a variable number of repeats, are best described as the re-activation of syllable-chains with auditory feedback, with the feedback strength reduced after each repetition (Wittenbach et al., 2015). In this work we focus on the non-repeat versions of the sequences, in which only the first syllable of any repetition is retained. For example, if the syllable sequence is $ABBBC$, the non-repeat version is ABC. In the rest of the paper, syllable sequences refer to the non-repeat versions.

Each syllable sequence is typically led by a variable number of introductory notes. These introductory notes are excluded in the analysis. All sequences have definite starts and ends. Thus the POMMs have two special states. One is the start state, from which all state transitions begin, and the other is the end state, at which all state transitions terminate. The POMMs are visualized with directed graphs (Fig. 1). Following the convention introduced previously (Jin and Kozhevnikov, 2011), we denote the start state as a pink oval marked with the symbol S. All other states are represented as ovals marked with associated syllables. The color of a state is cyan if it can transition to the end state, and is white otherwise. The end state is not shown in order to reduce clutter in the graph. State transitions are shown with arrows with transition probabilities written nearby. To reduce clutter, only transitions with probability $P > 0.01$ are shown.

Two types of context dependency

Context dependencies in syllable transitions can take two forms. In one form, certain transitions are prohibited depending on the context. A simple example is that the observed set contains two unique sequences: ACD and BCE, each with probability 0.5 (Fig. 1, Example 1). The transition $C \to D$ only occurs if C is preceded by A; and the transition $C \to E$ only occurs if C is preceded by B. In other words, sequences ACE and BCD are unobserved. We call this form the *type I context dependence.*

In the other form, context dependence manifests in the probabilities. A simple example modified from Example 1 is that the observed set contains sequences ACD, with probability 0.4; ACE, with probability 0.1; BCD, with probability 0.1; and BCE, with probability 0.4 (Fig. 1, Example 2). The transitions $C \to D$ and $C \to E$ are observed regardless of the preceding syllable; however, the transition probabilities are different when A precedes C than when B precedes C. We call this form the *type II context dependence.*

With the two examples we show that sufficient state multiplicity is required for capturing context dependencies in syllable transitions. For Example 1, consider constructing the Markov model for the set of observed sequences, which only requires calculating the transition probabilities between the syllables. The graph of the Markov model is shown in Fig. 1. The sequences can start with either syllable A or B with equal probability, hence the start state transitions to the states associated with syllables A or B (A-state or B-state) with probability 0.5. These two states transition to the C-state with probability 1. Since C can be followed by either D or E, the C-state transitions to the D-state or E-state with probability 0.5. From the start state, there are four possible state transition paths, generating four sequences ACD, ACE, BCD, and BCE, each with probability 0.25. Thus the Markov model overgeneralizes, creating unobserved sequences ACE and BCD.

To characterize the overgeneralization of a POMM, we introduce the concept of *sequence completeness* P_c, which is defined as the total probability of the POMM generating all unique sequences in the observed set:

$$P_c = \sum_{i=1}^{M} P_i,$$

where M is the number of unique sequences, and P_i is the probability of the i-th unique sequence. For Example 1, we have $P_c = 0.5$. The amount of overgeneralization is $1 - P_c$, which is the total probability of all unique sequences that the model can generate but are not in the observed set. The Markov model clearly does not capture the type I context dependence in the example. A more complex model has two states for syllable C, and the A-state and the B-state transition separately to these states (Fig. 1). This POMM generates two sequences ACD and BCE with probabilities 0.5 each, and $P_c = 1$ for the observed set.

Because P_c is the total probability of all unique sequences in the set, it is insensitive to the probabilities of individual unique sequences. Consider the Markov model for Example 2, which is the same as in Example 1 (Fig. 1). The Markov model generates all observed unique sequences, hence $P_c = 1$. Although the model does not overgeneralize, it does not capture the type II context dependence in Example 2. To reveal this deficiency, we need to compare probabilities of the unique sequences between the model and the observation.

A simple measure of the differences of the transition probabilities is the total variation distance (Gibbs and Su, 2002), defined as

$$d = \frac{1}{2} \sum_{i=1}^{M} |P_{i,o} - P_{i,m}|.$$

Here

$$P_{i,o} = \frac{N_i}{N}$$

is the observed probability of the i-th unique sequence, and is the ratio of the copy number N_i of this sequence in the observed set of N sequences; and $P_{i,m}$ is the normalized probability of the sequence computed with the model

$$P_{i,m} = \frac{P_i}{P_c}.$$

The normalization is to ensure that

$$\sum_{i=1}^{M} P_{i,m} = 1,$$

which is required since we are comparing $P_{i,m}$ to $P_{i,o}$, and $\sum_{i=1}^{M} P_{i,o} = 1$. For Example 2, the Markov model has $d = 0.3$. The model with two states for C, as shown in Fig. 1, can perfectly capture this type II context dependence with $d = 0$.

The total variation distance may not reveal type I context dependence. For Example 1, the Markov model generates the two observed sequences ACD and BCE with probabilities 0.25. However, after normalization the probabilities are 0.5. Hence we have $d = 0$ for the Markov model.

To capture both type I and type II context dependence, we combine P_c and d into a single measure

$$P_\beta = (1 - \beta)P_c + \beta(1 - d),$$

where β is the weight given to the total variation distance, and is a number between 0 and 1. We call this quantity the *augmented sequence completeness*. In this paper we set $\beta = 0.2$. We find that this choice gives a good balance in discovering both types of context dependencies in syllable transitions. A perfect model would have $P_\beta = 1$.

Neural correlates of state multiplicity

Within the framework of syllable-chains in HVC, it is natural to assume that the multiple states associated with one syllable correspond to multiple syllable-chains in HVC that drive the production of the same syllable (Jin, 2009; Cohen et al., 2020). In Example 1 discussed above, the POMM that fits the observed sequences has two states for syllable C. With two syllable-chains for C, the sequences ACD and ACB can be wired into two separate chains, as shown in Fig. 2a. This is the *intrinsic mechanism* for the state multiplicity in POMMs. This mechanism can account for the type II context dependence in Example 2 by introducing weaker connections from the end of the syllable-chain for C in ACD to the start of the syllable-chain for E, and from the end of the syllable-chain for C in BCD to the start of the syllable-chain for D, since the transition probabilities depend on the connection strength (Jin, 2009).

An alternative mechanism uses auditory feedback. In this case there is one syllable-chain for C, which connects to the syllable-chains for D and E. However, the activations of the syllable-chains for D and E are determined by the reafferent auditory inputs (Sakata and Brainard, 2006; Sakata and Brainard, 2008; Hanuschkin et al., 2011; Wittenbach et al., 2015). The auditory feedback from syllable A is sent to the syllable-chain for D; while the auditory feedback from syllable B is sent to the syllable-chain for E (Fig. 2b). The auditory inputs can bias the transitions from syllable-chain C to syllable-chains D and E (Jin, 2009; Hanuschkin et al., 2011; Wittenbach et al., 2015). With strong enough auditory inputs, the probability of transition from C to D should approach 1 when C is preceded by A. When C is preceded by B, the transition probability to D should approach 1. This is the *reafferent mechanism* for the state multiplicity.

These two mechanisms have different predictions for the effects of deafening. The intrinsic mechanism predicts that that the state multiplicity remains unchanged after deafening. The reafferent mechanism predicts that all state multiplicity disappears after deafening, and the song syntax will become Markovian. These predictions can be tested by inferring POMMs for the observed syllable sequences before and shortly after deafening.

Statistical test of POMM

To find the POMM that is compatible with the observed set of syllable sequences, we need to devise a way of statistically evaluating the validity of the POMM. This problem can be cast as hypothesis test, in which the null hypothesis is that the observed set is generated by the POMM. We can use P_β for this purpose. Ideally, P_β of the observed set computed with the

POMM should be 1, which indicates that the POMM generates all of the unique sequences in the observed set, and importantly, does not generate unobserved sequences; moreover, the probabilities of the unique sequences agree with the observations. In practice, due to the finite number N of sequences observed, it is possible that the observed set does not contain all possible sequences that the bird is capable of producing. Therefore, $P_c < 1$ could be due to the smallness of N, and not because the model overgeneralizes. Additionally, mismatch in the probabilities of the unique sequences could be due to the inaccurate measurements of the transition probabilities when N is finite.

To take into account the finite N effect, we generate random sets of N sequences from the POMM. For each generated set, we compute P_β with the POMM. The P_β distribution of the generated sets can be used to gauge the likelihood that the P_β of the observed set is drawn from the distribution. Specifically, we compute the probability p that the observed P_β is greater than the P_β of the generated sets. If $p < 0.05$, we conclude that the observed P_β is not likely drawn from the distribution, and the POMM is not likely the model that generates the observed set. If $p > 0.05$, the POMM is not statistically rejected and it is compatible with the observed set. In this work, we build the P_β distribution by generating 10000 random sets of N sequences from the POMM.

We illustrate this process with an example. In Fig. 3a, we show the “ground truth model”. It has 2 states for syllables A and C, and one state for each of syllables B, D, E. The model generates 7 unique sequences: A, probability 0.1; ACD, probability 0.36; ACE, probability 0.04; BCD, probability 0.05; BCE, probability 0.2; BAE, probability 0.125; and BA, probability 0.125. From the model, we generate three sets of “observed sequences” with $N = 10$, $N = 30$ and $N = 60$, as shown in the figure. Sequences generated from the ground truth model contain both type I and type II context-dependent syllable transitions.

We construct Markov models from the observed sets by computing the probabilities of starting or ending at each syllable, and the probabilities of transitioning from one syllable to another. The Markov models are shown in Fig. 3b. We generate 10000 random sets of N sequences from the Markov models, and compute P_β of these generated sets with the Markov model. The distributions of P_β are shown below the Markov models in Fig. 3b. The distributions shift towards 1 as N increases (Fig. 3b). We then calculate the P_β of the observed sets with the Markov models and indicate the values with red lines in Fig. 3b. The p-value is computed as the probability p that the observed P_β is greater than the P_β of the generated sets. In the examples shown in Fig. 3b, the p-values are $p = 0.12$ for $N = 10$; $p = 0.002$ for $N = 30$; and $p = 0$ for $N = 60$.

We run this process for 100 observed sets generated from the ground truth model for each N, and compute the p-value distributions. For $N = 10$, we find that $p = 0.27 \pm 0.28$; for $N = 30$, $p = 0.008 \pm 0.016$; and for $N = 60$, $p = 5 \times 10^{-6} \pm 3.2 \times 10^{-5}$. Therefore, for $N = 30$ and $N = 60$, the Markov model can be rejected based on the $p < 0.05$ criterion. For $N = 10$, however, the Markov model cannot be rejected, even though the ground truth model is non-Markovian.

If the ground truth model is Markovian, increasing N does not lead to rejection of the Markov model, as expected (supplementary Fig. S1). Although we used the Markov model as an example, this process of statistical testing based on P_β can be applied to any POMM.

Inferring POMM from observed sequences

Given a set of observed syllable sequences, we infer a POMM that is statistically compatible with the set. We also require that the POMM is a minimal model, such that the number of states for each syllable is as small as possible, and the transitions between the states are sparse. This is achieved through a procedure that consists of grid search in the state space, state reduction, and pruning of transitions between the states. We illustrate this procedure through the example shown in Fig. 3a.

A POMM is determined by the number of states for each syllable, and the transition probabilities between the states. All possible POMMs thus can be represented as grid points in the state space. For example, the grid point (1, 1, 1, 1, 1) represents the POMM with syllables A, B, C, D, E each having one state, which is the Markov model; and the grid point (2, 2, 2, 2, 2) represents the POMM with two states for each syllable. At each grid point, we find the transition probabilities between the states by maximizing the likelihood that the model generates the observed sequences using the Baum-Welch algorithm (Rabiner, 1989). To avoid local minima that the algorithm may encounter, we consider 100 runs of the algorithm with random seeds, and select the run with the largest likelihood.

The search starts with the Markov model, the grid point (1, 1, 1, 1, 1). The model is evaluated with the stopping criterion that it passes the P_β based statistical test with $p > 0.05$, as discussed above (Fig. 3). If the model does not satisfy the stopping criterion, the nearby grid points (2, 1, 1, 1, 1), (1, 2, 1, 1, 1), (1, 1, 2, 1, 1), (1, 1, 1, 2, 1), and (1, 1, 1, 1, 2) are accessed. Among them, the grid point with the largest likelihood is selected. If this newly selected point does not satisfy the stopping criterion, the search moves on to its nearby grid points. The process iterates until the stopping criterion is satisfied.

It is possible that the search ends up with a more complex POMM than needed because the path is guided by local information on the grid. We therefore test reducing the POMM by deleting states, which is the reverse process of the grid search. Specifically, for all syllables with multiple states, we delete one state for each. We select the deletion with the largest likelihood, and test whether the reduced POMM satisfies the stopping criterion. If the stopping criterion is satisfied, we go on to the next round of deletions. The process continues until the reduced POMM is rejected. The last deletion is then reversed.

After state reduction, we simplify the transitions between the states in the POMM. We systematically cut every transition and recalculate the maximum likelihood of the observed sequences. If the likelihood is larger than a threshold, the cut is accepted; otherwise the transition is retained. The threshold is set to the maximum likelihood of the POMM before cuts minus an estimate of the fluctuation of the likelihood due to inaccuracies in computing the likelihood, which is set to be the standard deviation of the likelihood in the 100 runs of the Baum-Welch algorithm with random seeds before the cuts. If the POMM after the accepted cuts no longer satisfy the stopping criterion, the threshold is raised and the cuts are redone.

We show the accuracy of the above procedure by inferring POMMs from 100 sets of N

observed sequences generated from the ground truth model (Fig. 3a). The results for $N = 10, 30, 90$ are shown in Fig. 4. We display typical POMMs inferred, and the distributions of the total number of states in the POMMs inferred from the 100 sets. For $N = 10$, the total number of states is mostly 5, and the Markov model is accepted. This is because for most sets of $N = 10$, the Markov model passes the statistical test (Fig. 3b). Some models have 4 states because syllables D or E may not appear in the observed sequences due to the small N. For $N = 30$, the total number of states ranges from 5 to 7. Typical POMMs with 6 states are shown in the figure. For $N = 90$, the total number of states is mostly 7, and the inferred POMMs have the same structure as the ground truth model.

This example shows that our procedure tends to fit a simpler POMM when the number of observed sequences is small. When the number is large, the procedure uncovers the ground truth model. Crucially, the procedure does not create more complex models than the ground truth model.

Effects of deafening on the POMM syntax of Bengalese finch songs

To see how auditory input affects the POMM syntax, we analyze songs of 6 adult Bengalese finches before and two days after deafening. The dataset was used previously for analyzing syllable repeats (Wittenbach et al., 2015). Here we focus on the non-repeat versions of the syllable sequences.

We first test if Markov models are statistically compatible with the observed syllable sequences using the $p > 0.05$ criterion. The results are shown in Fig. 5 for o10bk90, and in S2-S6 for the other five birds. Three birds have non-Markovian syntax before and after deafening (o10bk90, normal $p = 0$, deafened $p = 0$, Fig. 5; bfa16, normal $p = 0$, deafened $p = 0$, Fig. S3; o46bk78, normal $p = 0$, deafened $p = 0$, Fig. S6). The other three birds have non-Markovian syntax before deafening, but after deafening the Markovian syntax is not statistically rejected (bfa7, normal $p = 0$, deafened $p = 0.42$, Fig. S2; bfa14, normal $p = 0$, deafened $p = 0.56$, Fig. S5; bfa19, normal $p = 0.02$, deafened $p = 0.34$, Fig. S4). These results suggest that deafening reduces Bengalese finch song syntax from non-Markovian to Markovian for some birds but not for all.

Deafening also creates novel transitions between syllables, as well as novel starting and ending syllables. The transition probabilities of these novel transitions tend to be small (median $P = 0.04$), but 22% have probabilities larger than 0.1 (18 transitions out of 81). The majority of these novel transitions appear in two birds (27 for bfa14; 21 for bfa19). A small number (8) of transitions also disappear after deafening (median $P = 0.02$).

As observed in previous studies (Woolley and Rubel, 1997; Okanoya and Yamaguchi, 1997), deafening increases sequence variability. The variability of transitions from a given syllable i (or the start state) is quantified with the transition entropy as $S_i = -\sum_{j=1}^{M} p_{ij} \log_2 p_{ij}$, where M is the number of branches of the transitions, and p_{ij} is the probability of the j-th branch. If $M = 1$, the transition is stereotypical, and we have $S_i = 0$. For a given M, the entropy is maximum if the transition probabilities for all branches are equal. This maximum entropy increases with M. The median of transition entropies is significantly larger after deafening (median, 0.95, s.d., 0.55) than before (median, 0.35, s.d., 0.51; $p = 5 \times 10^{-6}$, Wilcoxon signed-rank one-sided test). The number of branches M is also significantly larger afer deafening (median, 4, s.d., 1.5) than before (median, 2, s.d., 0.90; $p = 9.8 \times 10^{-7}$, Wilcoxon signed-rank one-sided test).

We next construct POMMs from the observed syllable sequences before and after deafening. The inferred POMMs are shown in Figs. 6-11. In normal hearing condition, there are 44 syllables in the songs of the birds; among them, 25 require 1 state, 14 require 2 states, 2 require 3 states, and 3 require 4 states. So most syllables require 1 or 2 states. There are 77 states in the POMMs. Counting only transition branches with probabilities greater than 0.01, the majority of states have up to 3 outgoing branches (32, 29, 13 for branch numbers 1, 2, 3, respectively).

After deafening, there are 43 syllables (syllable g for bfa7 drops out after deafening). Most syllables (40) require only 1 state, and the remaining 3 require 2 states. There are 52 states in the POMMs. Counting only the transition branches with probability greater than 0.01, the branch numbers range from 1 to 7, with counts 2, 19, 7, 13, 6, 3, 2, respectively.

Deafening significantly reduces the state multiplicity, as measured by the number of extra states (defined as the number of states for the syllables minus the number of the syllables) (Fig. 12a, the Wilcoxon signed-rank one-sided test, $p = 0.016$). The mean normalized transition entropy between the states (transition entropy divided by $\log_2 M$, where M is the number of transitions from the state) is significantly larger after deafening for all but one bird (Wilcoxon signed-rank one-sided test, $p = 0.03$, tested with all birds). Thus, deafening reduces the complexity of song syntax, as indicated by the reduction of the extra number of states required. Additionally, transitions between the states become more random.

The POMMs reveal context dependencies in the syllable transitions. In the following, we show such dependencies for each bird before and after deafening. We first show the major syllable transitions in the observed sets. We then point out how reducing the state multiplicity by merging states associated with the same syllable makes the POMM overgeneralize or produce some subsequences with enhanced probabilities. This merging technique is inspired by the example shown in Fig. 1. The state-merged POMM retains all state transition branches of the original POMM, but the transition probabilities are re-calculated with the Baum-Welch algorithm using the sequences in the observed sets.

For each state-merged POMM, we use one or two selected subsequences for evaluation. We first calculate P_s of the subsequence in the observed set, defined as the fraction of sequences in the set that contain the subsequence. We then generate 10000 sets of N sequences from the POMM, where N is the number of sequences in the observed set. For each generated set, we compute P_s. This creates a distribution of P_s. We report the median value of P_s in this distribution to show how much the probability is enhanced. The significance of the enhancement is shown with the p-value, which is the probability p that P_s in the distribution is smaller than the observed P_s. The process is analogous to the test of POMMs shown in Fig. 3.

For o10bk90 in normal hearing condition, syllables f and g are represented by two states each, reflecting the following context dependence of syllable transitions (Fig. 6):

$$e \longrightarrow f_1 \longrightarrow g_1 \longrightarrow \square \ ,$$

$$\begin{tikzcd} & & a & \\ d \arrow[r] & f_2 \arrow[ur] \arrow[dr] & & \\ & & g_2 \arrow[r] & a \end{tikzcd} .$$

Here □ denote the end of the sequence, and the subscripts indicate different states for the same syllable. Merging f_1 and f_2 creates a subsequence

$$e \longrightarrow f \longrightarrow a \ .$$

This subsequence is unobserved, i.e. $P_s = 0$ in the observed set. From the distribution of P_s generated from the state-merged POMM we find that the median $P_s = 0.13$ and $p = 0.0001$, showing that the enhancement of P_s after merging the states is significant at the $\alpha = 0.05$ level. In the observed set, the subsequence

$$d \longrightarrow f \longrightarrow g \longrightarrow \square$$

is rare ($P_s = 0.016$). Merging g_1 and g_2 significantly increases the probability, with median $P_s = 0.27$ and $p = 0$.

After deafening, transition $S \to d$ is weakened, where S is the start state; and transitions $S \to a$ and $S \to g$ become stronger (Fig. 6). The state multiplicity for f persists, reflecting the context dependent transitions

$$e \longrightarrow f_1 \longrightarrow g \ ,$$

$$d \longrightarrow f_2 \longrightarrow a \ ,$$

which is the same as before deafening. As in normal hearing condition, merging f_1 and f_2 creates unobserved subsequencre

$$e \longrightarrow f \longrightarrow a$$

with median $P_s = 0.12$ and $p = 0$. The subsequence $d \to f \to g$ becomes rare after deafening ($P_s = 0.5$, before deafening; $P_s = 0.007$, deafened), indicating that deafening makes the transition $f_2 \to g_2$ rare. Syllable g is now represented with one state only, because this does not make the subsequence $d \to f \to g \to \square$ more frequent than observed, unlike in the normal hearing condition.

For bfa7 with normal hearing, syllable b has 2 states and syllables c and d have 4 states each (Fig. 7). The two states for b encode the following context dependence:

states merged	subsequence	P_s observed	median P_s	p
b_1, b_2	$abcdg$	0	0.07	0.004
c_1, c_2	$fcdg$	0	0.11	0
c_1, c_3	$bcdcdb$	0	0.04	0.0065
c_1, c_4	$fcdb$	0	0.07	0.0003
c_2, c_3	$abcdg$	0	0.04	0.0071
c_2, c_4	$bcdb$	0	0.04	0.007
c_3, c_4	$bcdb$	0	0.04	0.0078
d_1, d_2	$fcdg$	0	0.11	0
d_1, d_3	$bcdcdb$	0	0.04	0.0078
d_1, d_4	$fcdb$	0	0.07	0.0001
d_2, d_3	$abcdg$	0	0.04	0.0062
d_2, d_4	$bcdb$	0	0.04	0.0066
d_3, d_4	$bcdb$	0	0.04	0.0069

Table 1: Consequences of pairwise merging of states with the same syllables for bfa7 with normal hearing. Listed are the pair of states merged, subsequences examined, P_s of the subsequences in the observed set, median of the P_s distribution generated from the state-merged POMMs, and the p-value.

$$\begin{tikzcd} d \arrow[r] & b_1 \arrow[r] & c \arrow[r] & d \arrow[r] & g \end{tikzcd} ,$$

$$\begin{tikzcd} a \arrow[r] & b_2 \arrow[r] & c \arrow[r] & d \arrow[r] \arrow[l, bend right] & \square \end{tikzcd} .$$

The state multiplicity for c and d reflects the following context dependencies:

$$\begin{tikzcd} & & & & & a \\ f \arrow[r] & c_1 \arrow[r] & d_1 \arrow[r] & c_4 \arrow[r] & d_4 \arrow[ur] \arrow[r] \arrow[dr] \arrow[lll, bend right] & b \\ & & & & & \square \end{tikzcd} ,$$

$$\begin{tikzcd} d \arrow[r] & b \arrow[r] & c_2 \arrow[r] & d_2 \arrow[r] & g \end{tikzcd} ,$$

$$\begin{tikzcd} a \arrow[r] & b \arrow[r] & c_3 \arrow[r] & d_3 \arrow[r] \arrow[l, bend right] & \square \end{tikzcd} .$$

The consequences of merging states are summarized in Table 1.

Deafening leads to the appearance of $c \to a$ and $h \to a$ transitions, strengthening of $d \to a$ transition, and disappearance of $d \to c$ transition. Except for b, the sequence can now stop at all syllables. Interestingly, the $d \to g$ transition is lost and syllable g does not appear after deafening. The syntax is Markovian, suggesting that there is no context dependence.

For bfa16 in normal hearing condition, there are two states for syllables a, d, and e (Fig. 8). The two states for a encode the following context dependence:

$$\begin{tikzcd} S \arrow[r] & a_1 \arrow[r] & b \end{tikzcd} ,$$

$$\begin{tikzcd} & & \square \\ g \arrow[r] & a_2 \arrow[ur] \arrow[dr] & \\ & & d \end{tikzcd} .$$

Merging a_1 and a_2 creates an unobserved subsequence

$$\begin{tikzcd} S \arrow[r] & a \arrow[r] & d \end{tikzcd}$$

with median $P_s = 0.19$ and $p = 0$. The two states for d encodes the following context dependence:

$$\begin{tikzcd} c \arrow[r] & d_1 \arrow[r] & e \end{tikzcd} ,$$

$$\begin{tikzcd} a \arrow[r] & d_2 \arrow[r] & b \end{tikzcd} .$$

Merging d_1 and d_2 creates an unobserved subsequence

$$\begin{tikzcd} c \arrow[r] & d \arrow[r] & b \end{tikzcd}$$

with median $P_s = 0.22$ and $p = 0$. The two states for e encodes the context dependence

$$\begin{tikzcd} & & g \\ d \arrow[r] & e_1 \arrow[ur] \arrow[dr] & \\ & & f \end{tikzcd} ,$$

$$\begin{tikzcd} & & a_2 \\ f \arrow[r] & e_2 \arrow[ur] \arrow[dr] & \\ & & a_1 \end{tikzcd} .$$

Merging e_1 and e_2 creates unobserved subsequence

$$\begin{tikzcd} d \arrow[r] & e \arrow[r] & a \end{tikzcd}$$

with median $P_s = 0.38$ and $p = 0$.

The major effects of deafening are the loss of the transition $e_2 \to a_2$; the strengthening of the transition $e_2 \to a_1$; and the enhancement of stopping after g. The only state multiplicity left is for syllable e, which encodes the same context dependency as in the normal hearing condition. Merging the two states for e again creates unobserved subsequence $d \to e \to a$ with median $P_s = 0.05$ and $p = 0$.

For bfa19 in normal hearing condition, there are two states for syllables b, c, e, and f (Fig. 9). The state multiplicity for b and c encodes the context dependence

states merged	subsequence	P_s observed	median P_s	p
b_1, b_2	$abcf$	0	0.38	0.0001
c_1, c_2	$abcf$	0	0.38	0
e_1, e_2	$debcd$	0	0.29	0.0015
f_1, f_2	gfg	0	0.29	0.0011

Table 2: Consequences of pairwise merging of states with the same syllables for bfa19 with normal hearing.

$$\begin{tikzcd} & & & \square \\ a \arrow[r] & b_1 \arrow[r] & c_1 \arrow[ur] \arrow[dr] & \\ & & & d \end{tikzcd} ,$$

$$\begin{tikzcd} e \arrow[r] & b_2 \arrow[r] & c_2 \arrow[r] & f \end{tikzcd} .$$

The state multiplicity for e and f reflects the context dependency

$$\begin{tikzcd} d \arrow[r] & e_1 \arrow[r] & b \arrow[r] & c \arrow[r] & f_1 \arrow[r] & g \end{tikzcd} ,$$

$$\begin{tikzcd} & & & & & \mathrm{d} \\ g \arrow[r] & f_2 \arrow[r] & e_2 \arrow[r] & b \arrow[r] & c \arrow[ur] \arrow[dr] & \\ & & & & & \square \end{tikzcd} .$$

The consequences of pairwise state merging are shown in Table 2.

After deafening, many novel transitions appear, most notably $e \to g$ and $f \to g$ transitions. The model becomes Markovian, and all context dependencies disappear.

For bfa14 in normal hearing condition, the POMM has two states for c and g (Fig. 10), reflecting context dependence

$$\begin{tikzcd} b \arrow[r] & c_1 \arrow[r] & g_1 \arrow[r] & e \end{tikzcd} ,$$

$$\begin{tikzcd} & & & \square \\ f \arrow[r] & c_2 \arrow[r] & g_2 \arrow[ur] \arrow[dr] & \\ & & & e \end{tikzcd} .$$

Subsequence

$$\begin{tikzcd} b \arrow[r] & c \arrow[r] & g \arrow[r] & \square \end{tikzcd}$$

is rare in the observed sequences ($P_s = 0.03$). Merging c_1 and c_2 significantly boosts the probability, with median $P_s = 0.23$ and $p = 0$. Merging g_1 and g_2 does the same, with median $P_s = 0.08$ and $p = 0.013$.

For this bird, deafening creates numerous novel transitions with small probabilities (< 0.1)

(Fig. 10). Novel transitions with large probability (> 0.1) also occur, which include transitions $a \to h$, $b \to l$, $h \to f$, and $l \to g$, as well as from S to syllables b, c, e, f, l. Some transitions are weakened, which include transitions $l \to c$ and $f \to c$. The model becomes Markovian.

For o46bk78 with normal hearing, the song is described by a POMM with state multiplicity for multiple syllables (Fig. 11). There are 4 states for b, 3 states for a and c, and 2 states for d and e, respectively. The state multiplicity for syllable a reflects the following context dependencies:

$$\begin{tikzcd} & & & & & & \square \\ S \arrow[r] & a_1 \arrow[r] & b \arrow[l, bend right] \arrow[r] & c \arrow[r] & d \arrow[r] & e \arrow[ru] \arrow[rd] & \\ & & & & & & a \end{tikzcd},$$

$$\begin{tikzcd} & & & & & & & \square \\ b \arrow[r] & a_2 \arrow[r] & b \arrow[r] & c \arrow[r] & d \arrow[r] & e \arrow[r] & a \arrow[ru] \arrow[rd] & \\ & & & & & & & b \end{tikzcd},$$

$$\begin{tikzcd} & & \square \\ e \arrow[r] & a_3 \arrow[ru] \arrow[rd] & \\ & & b \end{tikzcd}.$$

The state multiplicity for syllable b reflects the following context dependencies:

$$\begin{tikzcd} & & & & & & \square \\ S \arrow[r] & a \arrow[r] & b_1 \arrow[l, bend right] \arrow[r] & c \arrow[r] & d \arrow[r] & e \arrow[ru] \arrow[rd] & \\ & & & & & & a \end{tikzcd},$$

$$\begin{tikzcd} & & c \arrow[r] & f \\ S \arrow[r] & b_2 \arrow[ru] \arrow[rd] & & \\ & & a \arrow[r] & b \end{tikzcd},$$

$$\begin{tikzcd} & & & & & & & & \square \\ S \arrow[r] & b \arrow[r] & a \arrow[r] & b_3 \arrow[r] & c \arrow[r] & d \arrow[r] & e \arrow[r] & a \arrow[ru] \arrow[rd] & \\ & & & & & & & & b \end{tikzcd},$$

$$\begin{tikzcd} e \arrow[r] & a \arrow[r] & b_4 \arrow[r] & c \arrow[r] & f \end{tikzcd}.$$

The state multiplicity for syllable c and d encodes the following context dependencies:

$$\begin{tikzcd} S \arrow[r] & a \arrow[r] & b \arrow[r] & c_1 \arrow[r] & d_1 \arrow[r] & e \arrow[r] & \square \end{tikzcd},$$

$$\begin{tikzcd} & & & & & & & & \square \\ & & a \arrow[r] & b \arrow[r] & c_2 \arrow[r] & d_2 \arrow[r] & e \arrow[r] & a \arrow[ru] \arrow[rd] & \\ S \arrow[r] & b \arrow[ru] \arrow[rd] & & & & & & & b \\ & & c_3 \arrow[r] & f & & & & & \end{tikzcd}.$$

The consequences of pairwise merging of states are shown in Table 3.

After deafening, a novel transition $d \to a$ appears. Moreover, the probability of stopping after syllable a is strongly enhanced. State multiplicity disappears except for syllable b, which

states merged	subsequence	P_s observed	median P_s	p
a_1, a_2	$Sabcdea\square$	0	0.04	0.029
a_1, a_3	$Sa\square$	0	0.06	0.014
a_2, a_3	$ba\square$	0	0.027	0
b_1, b_2	$Sbcd$	0.01	0.09	0.01
b_1, b_3	$Sabcdea\square$	0	0.04	0.03
b_1, b_4	$Sabcf$	0	0.07	0.0058
b_2, b_3	$babab$	0	0.14	0
b_2, b_4	$eaba$	0.03	0.17	0
b_3, b_4	$babcf$	0.03	0.23	0
c_1, c_2	$Sabcdea\square$	0	0.04	0.035
c_1, c_3	$Sabcf$	0	0.07	0.005
c_2, c_3	$babcf$	0	0.3	0
d_1, d_2	$Sabcdea\square$	0	0.04	0.025

Table 3: Consequences of pairwise merging of states with the same syllables for o46bk78 with normal hearing.

is still associated with two states, reflecting the context dependent transitions

$$\begin{tikzcd} S \arrow[r] & a \arrow[r] & b_1 \arrow[r] \arrow[l, bend right] & c \end{tikzcd} ,$$

$$\begin{tikzcd} & & c \\ S \arrow[r] & b_2 \arrow[ur] \arrow[dr] & \\ & & a \end{tikzcd} .$$

This is a type II context dependence. In both cases, syllable b is followed by syllable a or c. However, from b_1 the transition to c is favored, with probability 0.88; in contrast, from b_2 the transition to a is favored with probability 0.77. In the observed set, the subsequences

$$\begin{tikzcd} a \arrow[r] & b \arrow[r] & a \end{tikzcd} ,$$

$$\begin{tikzcd} S \arrow[r] & b \arrow[r] & c \end{tikzcd}$$

occur with probabilities $P_s = 0.21$ and $P_s = 0.19$, respectively. Merging b_1 and b_2 significantly enhances the probabilities, with median $P_s = 0.37$ and $p = 0.001$ for the first subsequence and with median $P_s = 0.56$ and $p = 0$ for the second subsequence.

Discussion

Deafening induces rapid changes in syllable sequences in Bengalese finch songs (Woolley and Rubel, 1997; Okanoya and Yamaguchi, 1997; Wittenbach et al., 2015). In this work we analyze the changes in song syntax by inferring minimal POMMs from syllable sequences. The multiplicity of states in the POMMs reveal context dependencies in syllable transitions (Jin, 2009; Jin and Kozhevnikov, 2011). We find that deafening reduces the state multiplicity but does not eliminate it. Our results indicate that intact auditory feedback plays an important but not exclusive role in creating context dependencies in Bengalese finch songs.

Previous deafening studies in the Bengalese finch emphasized the loss of sequence stereotypy shortly after deafening and suggested that online auditory feedback is required for producing stereotyped syllable sequences (Woolley and Rubel, 1997; Okanoya and Yamaguchi, 1997) . We confirm that deafening makes syllable sequences more random. However, an alternative explanation could be that the activity of the auditory system becomes more random after being deprived of inputs (Resnik and Polley, 2021). We find that deafening leads to the appearance of many novel transitions with small probabilities (< 0.1). Novel transitions with large probability also occur, but are less frequent. Some transitions with small probabilities disappear after deafening. The appearance (and disappearance) of transitions with small probabilities is consistent with the idea that the HVC activity is more random after deafening. NIf (the nucleus interfacialis of the nidopallium) is a major source of auditory inputs to HVC (Coleman and Mooney, 2004). During sleep, NIf activity drives random activations of HVC projection neurons (Hahnloser and Fee, 2007). It is conceivable that deafening deprives structured auditory inputs to NIf, and causes NIf to be randomly active during singing. Lesioning NIf in the Bengalese finch makes song sequences more stereotyped (Hosino and Okanoya, 2000), which suggests that NIf input is capable of influencing syllable transitions.

On average across the birds, the transition entropy at the branching points of syllable transitions tends to increase after deafening (Fig. 12b). This increase is mostly due to the branching points that have dominant transitions becoming more “equalized”, such that the branches have similar transition probabilities. Similar effects were seen in real-time manipulation of auditory feedback (Sakata and Brainard, 2006), cooling HVC (Zhang et al., 2017) or enhancing inhibition (Isola et al., 2020) in HVC of the Bengalese finch. It would be interesting to investigate whether there is a common neural mechanism across these manipulations.

In the framework of syllable-chains driving syllable productions (Hahnloser et al., 2002; Fee et al., 2004; Jin, 2009; Chang and Jin, 2009), transitions between syllable-chains are controlled by both the connections between the syllable-chains and the auditory inputs to the HVC projection neurons (Jin, 2009; Hanuschkin et al., 2011; Wittenbach et al., 2015). Strong auditory inputs can bias transitions towards the targeted branches. Context dependence can thus be encoded with many-to-one mapping from the syllable-chains to syllables (Jin, 2009; Cohen et al., 2020), or with the auditory feedback promoting different transitions depending on the preceding syllables (Fig. 2). These intrinsic and reafferent mechanisms can coexist. Deafening reduces context dependence, as indicated by the reduction of the state multiplicity in the POMMs after deafening. The state multiplicity remains for some syllables in some birds, suggesting the existence of the intrinsic mechanism. Additionally, because the delay of the auditory feedback is limited to about 70 - 90 ms (Sakata and Brainard, 2006), context dependence spanning many syllables is unlikely due to auditory feedback (Cohen et al., 2020). There are alternative frameworks on how syllables are driven by the song system in songbirds (Amador et al., 2013; Hamaguchi et al., 2016; Troyer et al., 2017). It would be interesting to show how these frameworks can explain our observations on the context-dependent syllable transitions in Bengalese finch songs.

Our method of inferring a POMM from observed sequences is conservative. The method is designed to find the minimal POMM given the observed sequences. When the number of observed sequences is small, the method tends to underestimate the true number of multiple states (Fig. 4). This is because not all context dependencies are sufficiently represented in the observed sequences. One way to gauge whether there are enough number of observed sequences is to see if the sequence completeness P_c computed with the POMM is close to 1 for the observed sequences. The quantity $1 - P_c$ can be used as a rough estimate of the total probability of the missing unique sequences.

We identify two types of context dependencies. Simple models that are incapable of capturing type I context dependencies overgeneralize, creating unobserved sequences. This is captured with P_c. In the case of sufficient number N of observed sequences, $1 - P_c$ is the total probability of the unobserved sequences. A perfect model should have $P_c = 1$. However, a model could have $P_c = 1$ but still miss type II context dependencies, which describe how transition probabilities change depending on the preceding syllables. This type can be captured by the total variation distance d, which is the sum of the differences of the model's and the observed probabilities of the unique sequences in the observed set. To capture both types of context dependencies, we combine P_c and d with a parameter β into the augmented sequence completeness P_β. An ideal model should have $P_\beta = 1$. Accurate measurements of the sequence probabilities require large N. If N is small, type II context dependencies may be obscured by the fluctuations in the measured probabilities. In this case it is better to de-emphasize the contribution of d by setting β close to 0. We find that setting $\beta = 0.2$ is a reasonable choice for our data set. Because P_c is the sum of the probabilities of the unique sequences, it is more robust against inaccurate measurements of the probabilities.

The method depends on the distribution of P_β for the sequences sampled from the candidate POMM. Some sequences that the model generates may be not observed not because the model overgeneralizes, but because there is not enough number of observations. This finite N effect can be estimated by sampling sets of N sequences from the model and computing P_β. This distribution is used to calculate the p-value of the P_β of the observed set computed with the POMM. We used the criteria $p < 0.05$ for rejecting the POMM. Lowering this cut off value so that rejection is more stringent should enable acceptance of POMMs with fewer number of extra states. Our approach for deriving POMM from observed sequences is computationally intensive. The major cost is the sampling step. It would be interesting to investigate better methods for estimating the state multiplicity. One possibility is to measure the predictive information in the syllable sequences and infer the number of parameters needed for encoding the sequence complexity (Bialek et al., 2001).

POMMs were inferred in a previous study by fitting probabilities distributions such as N-gram distributions, which are the probabilities of subsequences of length N (Jin and Kozhevnikov, 2011). The method involved multiple heuristic steps, and was not easy to implement. Additionally, the method required a large N because it relied on accurate measurements of the probabilities. In contrast, our method is principled, and can work with smaller N. Even though our method does not directly fit N-grams, the statistics of 2- to 7-grams agree between the observed sequences and the sequences generated by the POMMs (Fig. S7 and Fig. S8).

In conclusion, we devised a method of inferring minimal POMMs from observed sequences. Application of the method to the syllable sequences of Bengalese finch songs before and after deafening suggests that the auditory system helps to create context-dependences in syllable transitions. Our method should be broadly applicable to other animal behavioral sequences.

Materials and Methods

Data set

The data set in this work was previously used for analyzing syllable repeats (Wittenbach et al., 2015) (available for download from http://personal.psu.edu/dzj2/SharedData/KrisBouchard/). Details of recording songs, annotating syllables, and deafening through bilateral cochlear removal, as well as the Ethics Statement can be found in the published paper (Wittenbach et al., 2015). We specifically used the data collected from six male adult Bengalese finches before and after deafening (labeled bfa14, bfa16, bfa19,bfa7, o10bk90, and o46bk78).

In the data set, syllables are labelled a through l, and x through z. Some ambiguous syllables are noted with symbols 0 and $-$, and they are skipped. Bengalese finch song bouts typically begin with short introductory notes. They are labeled as i, j and k. We define song sequences as segments of syllables that are bracketed by periods of introductory notes and the end of the recordings.

POMM

A POMM is specified by a state vector $V = [S, E, s_3, s_4, \cdots, s_n]$, where $s_1 = S$ and $s_2 = E$ are the start and the end states, n is the total number of states, and s_i for $i = 3, \cdots, n$ is the syllable symbol associated with the ith state. The same syllable symbol can appear multiple times in the state vector. Transitions between the states are described by a transition matrix T, whose element T_{ij} gives the probability of transition from state i to state j. There are no transitions to the start state, i.e. $T_{i1} = 0$; and there are no transitions from the end state, i.e. $T_{2j} = 0$.

Sequence generation from a POMM starts with the S state. At state i, the next state j is chosen with the probabilities T_{ij} among possible choices of state 2 to state n. Once chosen, the symbol s_j is added to the sequence. This process repeats until the E state is reached, at which point the sequence generation is complete.

A POMM is visualized with the software Graphviz (Ellson et al., 2001). To reduce clutter, only transitions with probabilities larger than 0.01 are shown. Additionally, the E state is not shown. Instead, the states that can transition to the E state are shown in cyan. The transition probability from one state to the E state is 1 minus the sum of the transition probabilities to other states. If a state does not transition to the E state with a probability larger than 0.01, the state is shown as white. The start state is shown in pink.

Markov model

A Markov model is a special case of POMM for which each syllable symbol appears only once in the state vector. The transition probabilities T can be computed as

$$T_{ij} = \frac{N_{ij}}{N_i},$$

where N_i is the total number of times s_i appears in the set Y of sequences, and N_{ij} is the total number of the times that the two-symbol subsequence $s_i s_j$ appears in Y. Note that

$$N_i = \sum_{j=1}^{n} N_{ij},$$

so we only need to compute N_{ij}.

Baum-Welch algorithm

Computing T for POMM with state multiplicity is more complicated than that for the Markov model, but the approach is similar. Starting from a set of random transition probabilities, the state transition sequences that correspond to the syllable sequences in Y are worked out. The transition probabilities are then updated according to

$$T_{ij} = \frac{N_{ij}}{N_i},$$

where N_i is number of times the state i appears in the state sequences, and N_{ij} is the number of times the subsequence of states ij appears. With the updated T, the process is repeated. The process stops when the changes in T is smaller than 10^{-6}. Because the result might be dependent on the initialization of T, the process is run for 100 times with different seeds for random number generator. The T that maximizes the probabilities of generating Y from the POMM is selected.

The computation is efficiently implemented with the Baum-Welch algorithm (Rabiner, 1989). Consider a sequence $y_1 y_2 \cdots y_t \cdots y_m$ in the set Y. Here t is the step in the sequence and m is the maximum length of the sequence. The algorithm consists of three parts. First, calculate the forward probability $\alpha_i(t)$, which is the probability of being at state i at step t given the proceeding sequence is $y_1 y_2 \cdots y_{t-1}$. This is computed iteratively with

$$\alpha_i(t+1) = \delta_i(y_{t+1}) \sum_{j=1}^{n} \alpha_j(t) T_{ji}.$$

Since all sequences start from the S state, the initial condition is $\alpha_1(0) = 1$ and $\alpha_j(0) = 0$ for all $j \neq 1$. Here $\delta_i(y_{t+1}) = 1$ if the symbol y_{t+1} at step $t+1$ is the same as the symbol s_i associated with state i; otherwise, $\delta_i(y_{t+1}) = 0$. Second, calculate the backward probability $\beta_i(t)$, which is the probability being at state i at step t and the follow-up sequence is $y_{t+1}, \cdots, y_m$. This is calculated iteratively with

$$\beta_i(t) = \delta_i(y_t) \sum_{j=1}^{n} T_{ij} \beta_j(t+1).$$

Since all sequences end at the end state, the initial condition is $\beta_2(m+1) = 1$ and $\beta_j(m+1) = 0$ for all $j \neq 2$. Third, calculate N_i and N_{ij}. The forward and backward probabilities $\alpha_i(t)$ and $\beta_i(t)$ should be computed for each sequence in Y. The number of transition from state i to state j is given by

$$N_{ij} = \sum_{Y} \sum_{t=1}^{m} \alpha_i(t) T_{ij} \beta_j(t+1).$$

For a given sequence $y_1 y_2 \cdots y_m$, the probability that the POMM generates it is given by

$$P_y = \alpha_2(m+1),$$

which is the forward probability of ending at the end state at step $m+1$.

The total probability of the set Y is given by

$$P_Y = \Pi_{y \in Y} P_y.$$

It is most convenient to use the log likelihood, which is

$$L_Y = \log P_Y = \sum_{y \in Y} \log P_y.$$

Sequence completeness, total variation distance and augmented sequence completeness

For a set of sequences Y, the sequence completeness on a POMM is computed as

$$P_c = \sum_{y \in Y} P_y,$$

where y is a unique sequence in Y. The sum is over all the unique sequences in the set.

For a set of observed sequences Y_o, the total variation distance is defined as

$$d = \frac{1}{2} \sum_{y \in Y_o} |P_y - P_{y,m}|.$$

Here $P_{y,m}$ is the probability of the unique sequence y computed on the POMM and then *normalized* among the unique sequences such that

$$\sum_{y \in Y_o} P_{y,m} = 1.$$

This normalization is necessary because $P_{y,m}$ is compared to P_y, which is normalized:

$$\sum_{y \in Y_o} P_y = 1.$$

The total variation distance ranges from 0 to 1.

The augmented sequence completeness is defined as

$$P_\beta = (1 - \beta) P_c + \beta (1 - d).$$

Here β is a parameter that can be chosen in the range $(0, 1)$. The value of P_β ranges from 0 to 1. A perfect POMM for the observed set should yield $P_\beta = 1$ because $P_c = 1$ and $d = 0$. When N is small, the measurements of P_y are not accurate. For this case, the contribution from d

should be reduced by taking a small value for β. In our work, we chose $\beta = 0.2$.

Statistical test

To test whether an observed set Y_o with N sequences could be generated from a POMM, we generate $M = 10000$ sets of N sequences, and compute the P_β of the generated sets, which gives a distribution of P_β. We also compute the augmented sequence completeness $P_{\beta,o}$ of the observed set. In the distribution, we count the number K of P_β that are smaller or equal to $P_{c,o}$. To avoid small fluctuations in P_β making K artificially small, we added 10^{-10} to $P_{\beta,o}$. The p-value is

$$p = \frac{K}{M}.$$

The POMM is rejected if $p < 0.05$, and accepted otherwise.

Inferring minimal POMM

For a given set Y of N syllable sequences, the minimal POMM is inferred through three steps: grid search in the state space; state deletion; and removal of transitions. Let k be the number of syllables. The grid space has k dimensions, and the grid points $(x_1, x_2, \cdots, x_k)$ specifies a state vector V in which syllable s_i appears x_i times. Grid search starts with the Markov model $(1, 1, \cdots, 1)$. The model is tested for statistical significance of the P_β of Y on the model. If the Markov model is rejected, the nearby grid points $(2, 1, \cdots, 1), (1, 2, \cdots, 1), \cdots, (1, 1, \cdots, 2)$ are evaluated. The transition matrix T for each corresponding POMM is inferred using the Baum-Welch algorithm. The grid point with the maximum log-likelihood is selected, and the corresponding POMM is tested for the P_β significance. If rejected, the nearby points of the newly selected grid point are evaluated. This process continues, until one POMM is accepted according to the P_β statistical test.

Because grid search is a local "hill climbing" scheme, the POMM at which the search stops may not be the minimal POMM. We therefore perform state deletion, which is opposite of grid search. From the accepted POMM $(x_1, x_2, \cdots, x_k)$ in the grid search, we test grid points with one less number of states for one of the syllables: $(x_1 - 1, x_2, \cdots, x_k), (x_1, x_2 - 1, \cdots, x_k), (x_1, x_2, \cdots, x_k - 1)$. The grid point with the maximum log-likelihood is selected, and the POMM is tested for the P_β statistics. If the POMM is accepted, the next round of state deletion is performed. This process repeats, until no POMM at the tested grid points is accepted. The last accepted POMM in the process is the minimal POMM.

The final step is minimization of the number of transitions in the POMM. We first remove all transitions with probability smaller than 0.001. We then remove the remaining transitions one by one, and re-compute the transition matrix T after each removal. To remove a transition from state i to state j, we set $T_{ij} = 0$ in the initial transition matrix for the Baum-Welch algorithm. The algorithm ensures that this transition element remains 0. If the log-likelihood remains within the threshold, the removal is accepted and kept; otherwise the removal is rejected and reversed. The threshold is

$$L_\theta = L_{max} - \mu\sigma_L,$$

where L_{max} is the log-likelihood of the original POMM before any deletions, and σ_L is the standard deviation of the log-likelihood of the 100 runs of Baum-Welch algorithms with different random seeds. The parameter μ is set to 1. If after the deletions the p-value of the P_β test goes below 0.05, μ is reduced to 0.5, and the deletion process is done again. This reduction in μ is rarely needed.

Probability of finding a subsequence

The probability P_s of finding a subsequence in a set Y is defined as

$$P_s = \frac{K}{N},$$

where N is the number of sequences in the set, and K is number of sequences that contains the subsequence.

State merging tests

To evaluate the context dependent syllable transitions encoded by state multiplicity in a POMM, we perform pairwise state merging tests. The merged state retains all transitions to and from the two states. The transition probabilities of the state-merged POMM are recomputed using the Baum-Welch algorithm and the observed set Y_o. By examining the states transitioning into the two states, and the states that follow the two states, we find possible subsequences that can show overgeneralization after the state merger. We find a subsequence that either is unseen in the observed set ($P_{s,o} = 0$) or has small probability $P_{s,o}$. To see whether the subsequence is significantly more probable in the sequences generated from the state-merged POMM, we generate 10000 sets of N sequences from the POMM. Here N is the number of sequences in Y_o. For each generated set, we compute P_s. This creates a distribution. We count the number of P_s that is smaller than or equal to $P_{s,o} + 10^{-10}$. The p-value is the ratio of this number and 10000. We add a small number 10^{-10} to $P_{s,o}$. This is for avoiding artificially lowering p-value due to those P_s that are equal to $P_{s,o}$. For example, if the subsequence is unobserved ($P_{s,o} = 0$) and the state-merged POMM does not generate it either, we would have a situation that $P_s = 0$ for all of the sampled set. By adding the small number to $P_{s,o}$, we ensure that $p = 1$, as it should be. If $p < 0.05$, we conclude that the enhancement of P_s after state merger is significant.

Wilcoxon signed-rank test

For comparing distributions of paired data in Fig. **??**, we use Wilcoxon signed-rank test using scipy.stats.wilcoxon, which is in the Python module scipy.

References

Amador A, Perl YS, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495:59.

Bialek W, Nemenman I, Tishby N (2001) Predictability, complexity, and learning. *Neural computation* 13:2409–2463.

Chang W, Jin DZ (2009) Spike propagation in driven chain networks with dominant global inhibition. *Physical Review E* 79:051917.

Cohen Y, Shen J, Semu D, Leman DP, Liberti WA, Perkins LN, Liberti DC, Kotton DN, Gardner TJ (2020) Hidden neural states underlie canary song syntax. *Nature* 582:539–544.

Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. *Journal of Neuroscience* 24:7251–7265.

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience* 22:567–631.

Egger R, Tupikov Y, Elmaleh M, Katlowitz KA, Benezra SE, Picardo MA, Moll F, Kornfeld J, Jin DZ, Long MA (2020) Local axonal conduction shapes the spatiotemporal properties of neural sequences. *Cell* 183:537–548.

Ellson J, Gansner E, Koutsofios L, North SC, Woodhull G (2001) Graphviz – open source graph drawing tools In *International Symposium on Graph Drawing*, pp. 483–484. Springer.

Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. *Annals of the New York Academy of Sciences* 1016:153–170.

Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. *International statistical review* 70:419–435.

Hahnloser RH, Fee MS (2007) Sleep-related spike bursts in hvc are driven by the nucleus interface of the nidopallium. *Journal of neurophysiology* 97:423–435.

Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65.

Hamaguchi K, Tanaka M, Mooney R (2016) A distributed recurrent network contributes to temporally precise vocalizations. *Neuron* 91:680–693.

Hanuschkin A, Diesmann M, Morrison A (2011) A reafferent and feed-forward model of song syntax generation in the bengalese finch. *Journal of computational neuroscience* 31:509–532.

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

Isola GR, Vochin A, Sakata JT (2020) Manipulations of inhibition in cortical circuitry differentially affect spectral and temporal features of bengalese finch song. *Journal of Neurophysiology* 123:815–830.

Jin DZ (2009) Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus hvc. *Physical Review E* 80:051902.

Jin DZ (2013) The neural basis of birdsong syntax. *Progress in cognitive science: From cellular mechanisms to computational theories* .

Jin DZ, Kozhevnikov AA (2011) A compact statistical model of the song syntax in bengalese finch. *PLoS computational biology* 7:e1001108.

Jin DZ, Ramazanoğlu FM, Seung HS (2007) Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area hvc. *Journal of computational neuroscience* 23:283–299.

Jun JK, Jin DZ (2007) Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS One* 2:e723.

Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. *Nature* 468:394.

Lynch GF, Okubo TS, Hanuschkin A, Hahnloser RH, Fee MS (2016) Rhythmic continuous-time coding in the songbird analog of vocal motor cortex. *Neuron* 90:877–892.

Markowitz JE, Ivie E, Kligler L, Gardner TJ (2013) Long-range order in canary song. *PLoS computational biology* 9:e1003052.

Okanoya K (2004) The bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences* 1016:724–735.

Okanoya K, Yamaguchi A (1997) Adult bengalese finches (lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. *Journal of neurobiology* 33:343–356.

Picardo MA, Merel J, Katlowitz KA, Vallentin D, Okobi DE, Benezra SE, Clary RC, Pnevmatikakis EA, Paninski L, Long MA (2016) Population-level representation of a temporal sequence underlying song production in the zebra finch. *Neuron* 90:866–876.

Rabiner LR (1989) A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77:257–286.

Resnik J, Polley DB (2021) Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron* 109:984–996.

Sakata JT, Brainard MS (2006) Real-time contributions of auditory feedback to avian vocal motor control. *Journal of Neuroscience* 26:9619–9628.

Sakata JT, Brainard MS (2008) Online contributions of auditory feedback to neural activity in avian song control circuitry. *Journal of Neuroscience* 28:11378–11390.

Troyer TW, Brainard MS, Bouchard KE (2017) Timing during transitions in bengalese finch song: implications for motor sequencing. *Journal of neurophysiology* 118:1556–1566.

Tupikov Y, Jin DZ (2021) Addition of new neurons and the emergence of a local neural circuit for precise timing. *PLoS computational biology* 17:e1008824.

Wittenbach JD, Bouchard KE, Brainard MS, Jin DZ (2015) An adapting auditory-motor feedback loop can contribute to generating vocal repetition. *PLoS computational biology* 11:e1004471.

Woolley SM, Rubel EW (1997) Bengalese finches lonchura striata domestica depend upon auditory feedback for the maintenance of adult song. *Journal of Neuroscience* 17:6380–6390.

Woolley SM, Rubel EW (2002) Vocal memory and learning in adult bengalese finches with regenerated hair cells. *Journal of Neuroscience* 22:7774–7787.

Zhang YS, Wittenbach JD, Jin DZ, Kozhevnikov AA (2017) Temperature manipulation in songbird brain implicates the premotor nucleus hvc in birdsong syntax. *Journal of Neuroscience* 37:2600–2611.

Figure Legends

Figure 1: **Two types of context dependent syllable transitions**. Two examples are used to illustrate the computations of sequence completeness P_c, the total variation distance d, and the augmented sequence completeness P_β. Example 1 shows type I context dependence, and Example 2 shows type II context dependence.

Figure 2: **Neural mechanisms of POMM**. Schematics of how chain networks in HVC can be wired to implement the state multiplicity in POMMs. **a**. In the intrinsic mechanism, the multiple states for a syllable (C in this example) correspond to multiple syllable-chains that drive the production of the same syllable. **b**. In the re-afferent mechanism, the multiple states are due to auditory feedback biasing transition probabilities at the branching points.

Figure 3: **Statistical test of a POMM**. **a**. The ground truth POMM for generating the "observed set" of sequences from which the Markov models are derived. The POMM has two states for syllables A and C, and one states for syllables B, D, and E. The two states for A encodes type I context dependence, and the two states for C encodes type II context dependence. The sequences generated from the POMM are shown for $N = 10, 30, 60$. **b**. Markov models derived from the observed sets (up) and the distributions of P_β of 10000 sets of N sequences generated from the Markov models. The redlines indicate the P_β of the generated sequences computed with the Markov model. Three cases for $N = 10, 30, 60$ are shown.

Figure 4: **Derived POMMs for the example**. POMMs are derived from 100 sets of $N = 10, 30, 90$ generated from the ground truth model shown in Fig. 3a. Typical structures of the POMMs (top) and distributions of the number of states for the syllables (bottom) are shown.

Figure 5: **Test of Markov model for bird o10bk90**. The Markov models (top) and the P_β distributions (bottom) are shown for the normal hearing condition (left) and after deafening (right). The red lines are P_β of the observed sets computed with the Markov models. For both before and after deafening, the Markov models are rejected ($p = 0$ in both cases).

Figure 6: **POMM for bird o10bk90**. The POMMs before (left) and after (right) deafening are shown. The syllables with multiple states are highlighted with red. The p-values, the number N of sequences in the observed sets, and the P_β are displayed. Before deafening, syllables f and g each have two states. After deafening, f still has two states but g has one state.

Figure 7: **POMM for bird bfa7** . Same as in Fig. 6. Before deafening, syllable b has 2 states, and syllables c and d each has 4 states. After deafening, there is no state multiplicity. Note that syllable g is dropped after deafening.

Figure 8: **POMM for bird bfa16**. Same as in Fig. 6. Before deafening, syllables a, d and e each has 2 states. After deafening, only syllable e retains 2 states.

Figure 9: **POMM for bird bfa19**. Same as in Fig. 6. Before deafening, syllables b, c, e and f each has 2 states. After deafening, the state multiplicity disappears. Many novel transitions appear after deafening for this bird.

Figure 10: **POMM for bird bfa14**. Same as in Fig. 6. Before deafening, syllables c and g each has 2 states. After deafening, the state multiplicity disappears. Many novel transitions appear after deafening for this bird.

Figure 11: **POMM for bird o46bk78**. Same as in Fig. 6. Before deafening, all but one syllable f has multiple states (a, 3; b, 4; c, 3; d, 2; and e 2). After deafening the many-to-one disappears for all but syllable b, which still has 2 states.

Figure 12: **Summary of the effects of deafening on POMM**. (Left) The total numbers of extra states in POMMs decrease for all birds. (Right) The mean normalized transition entropies at branching points in the POMMs increase for all but one bird (bfa16), indicating that the transitions at branching points tend to become equally probable after deafening.

Figure S1: **(Supplementary) Statistical test of Markov model**. **a**. The ground truth model is a Markov model. Contrast this with the model in Fig. 3a. **b**. Examples of sequences generated from the ground truth model. **c**. From the "observed" sets of N sequences generated with the ground truth model ($N = 10, 30, 60$), Markov models are derived. The Markov models are tested with the distribution of P_β. The red lines indicate the P_β of the generated sequences from the Markov models. As expected, for all N the Markov model is not rejected.

Figure S2: **(Supplementary) Test of Markov model for bird bfa7**. Same as in Fig. 5. Before deafening, the Markov model is rejected ($p = 0$). After deafening, the Markov model is not rejected ($p = 0.42$).

Figure S3: **(Supplementary) Test of Markov model for bird bfa16**. Same as in Fig. 5. Both before and after deafening, the Markov models are rejected ($p = 0$ in both cases).

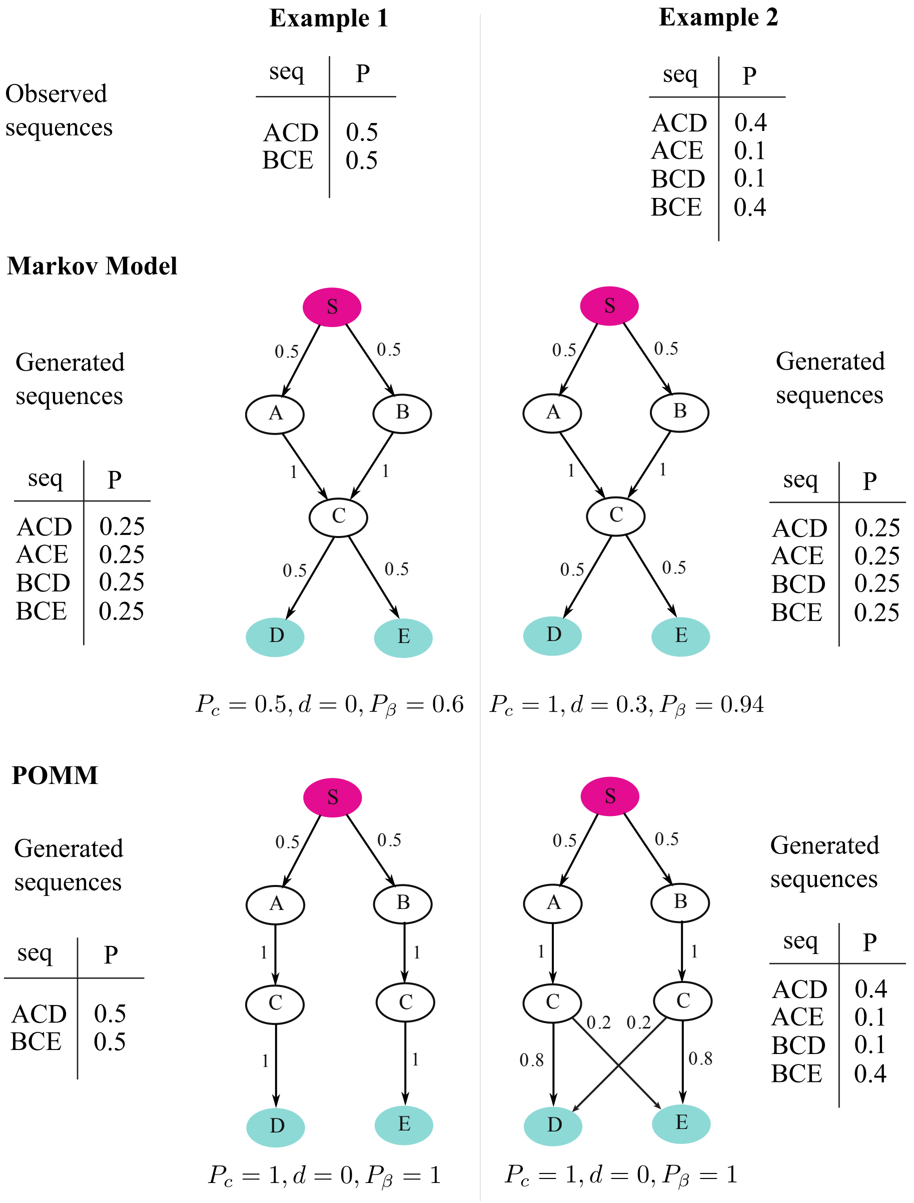
Figure S4: **(Supplementary) Test of Markov model for bird bfa19**. Same as in Fig. 5. Before deafening, the Markov model is rejected ($p = 0.02$). After deafening, the Markov model is not rejected ($p = 0.34$).

Figure S5: **(Supplementary) Test of Markov model for bird bfa14**. Same as in Fig. 5. Before deafening, the Markov model is rejected ($p = 0$). After deafening, the Markov model is not rejected ($p = 0.56$).

Figure S6: **(Supplementary) Test of Markov model for bird o46bk78**. Same as in Fig. 5. Both before and after deafening, the Markov models are rejected ($p = 0$ in both cases).

Figure S7: **(Supplementary) Comparisons of N-gram distributions in normal hearing condition**. For each bird, the probability distributions of 2- to 7 -grams of the sequences in the observed set are plotted in red. The N-grams are sorted in the decreasing orders of probabilities in the red curves. For comparisons, the probabilities of the same N-grams are computed for 100 sets of sequences generated from the POMM. Each set contains the same number of sequences as in the observed set. The N-gram probabilities are plotted with gray lines. For all birds, the red lines overlap with the gray lines, suggesting that the N-gram distributions agree between the observed sets and the generated sets.

Figure S8: **(Supplementary) Comparisons of N-gram distributions after deafeninig**. The same as in Fig. S7 but for the deafened cases.



Example 1, observed sequences:

seq	P
ACD	0.5
BCE	0.5

Example 2, observed sequences:

seq	P
ACD	0.4
ACE	0.1
BCD	0.1
BCE	0.4

Markov Model, Example 1, generated sequences:

seq	P
ACD	0.25
ACE	0.25
BCD	0.25
BCE	0.25

$P_c = 0.5, d = 0, P_\beta = 0.6$

Markov Model, Example 2, generated sequences:

seq	P
ACD	0.25
ACE	0.25
BCD	0.25
BCE	0.25

$P_c = 1, d = 0.3, P_\beta = 0.94$

POMM, Example 1, generated sequences:

seq	P
ACD	0.5
BCE	0.5

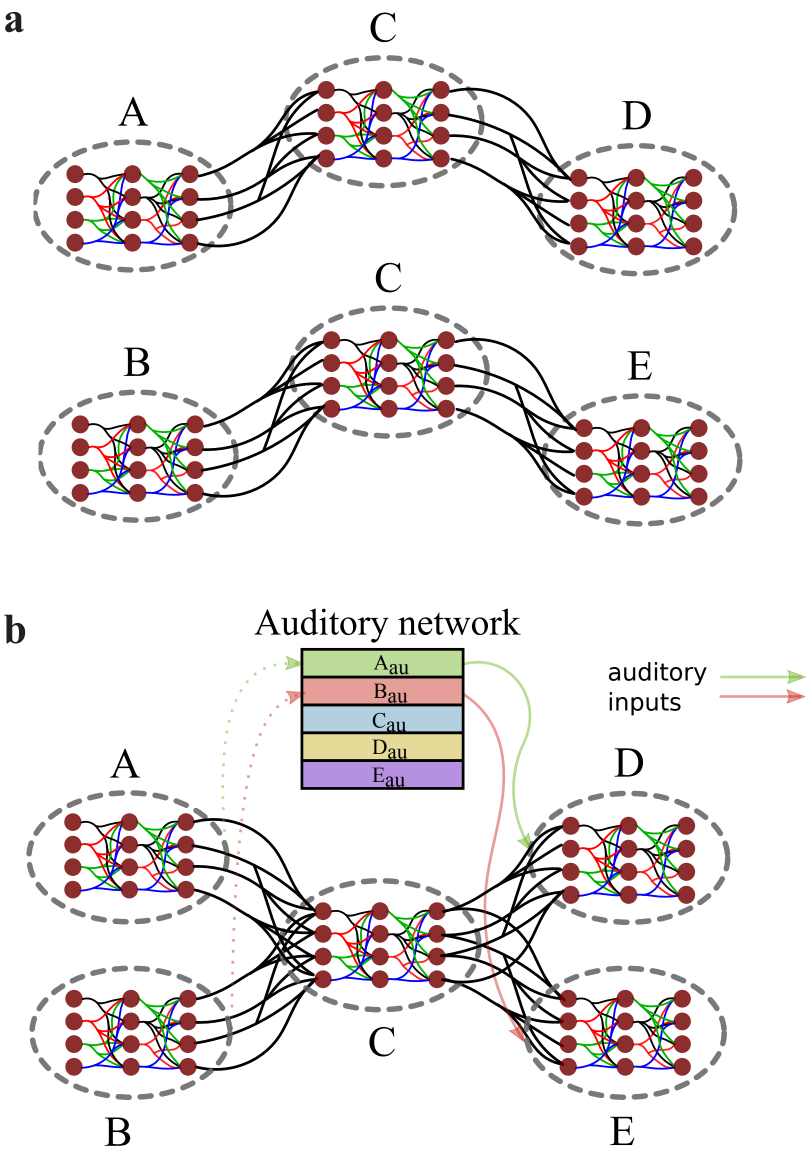
$P_c = 1, d = 0, P_\beta = 1$

POMM, Example 2, generated sequences:

seq	P
ACD	0.4
ACE	0.1
BCD	0.1
BCE	0.4

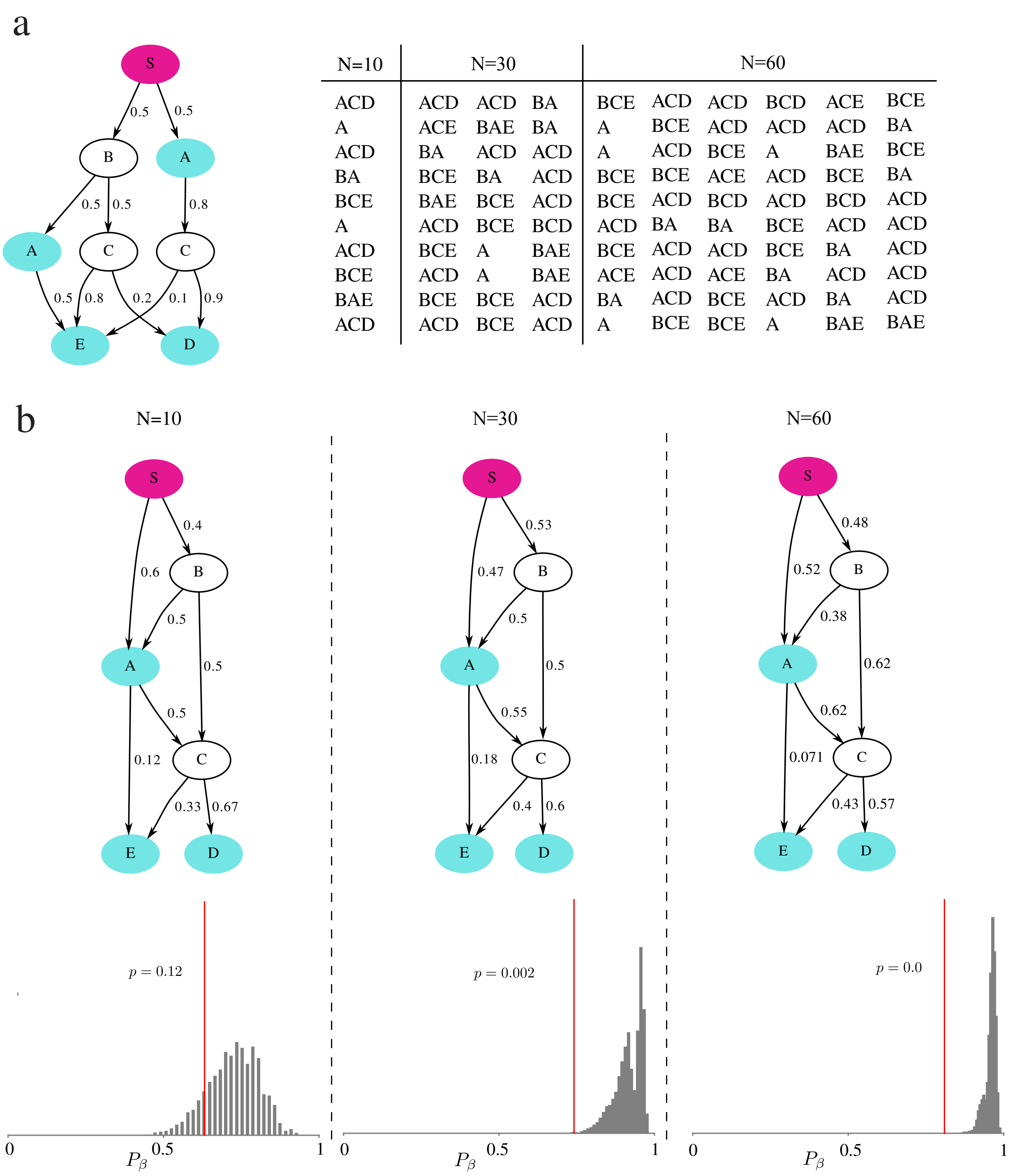
$P_c = 1, d = 0, P_\beta = 1$

Fig. 1 Two types of context dependent syllable transitions.



739

740 **Fig. 2 Neural mechanisms of POMM.**



N=10	N=30			N=60					
ACD	ACD	ACD	BA	BCE	ACD	ACD	BCD	ACE	BCE
A	ACE	BAE	BA	A	BCE	ACD	ACD	ACD	BA
ACD	BA	ACD	ACD	A	ACD	BCE	A	BAE	BCE
BA	BCE	BA	ACD	BCE	BCE	ACE	ACD	BCE	BA
BCE	BAE	BCE	ACD	BCE	ACD	BCD	ACD	BCD	ACD
A	ACD	BCE	BCD	ACD	BA	BA	BCE	ACD	ACD
ACD	BCE	A	BAE	BCE	ACD	ACD	BCE	BA	ACD
BCE	ACD	A	BAE	ACE	ACD	ACE	BA	ACD	ACD
BAE	BCE	BCE	ACD	BA	ACD	BCE	ACD	BA	ACD
ACD	ACD	BCE	ACD	A	BCE	BCE	A	BAE	BAE

Fig. 3 Statistical test of a POMM.

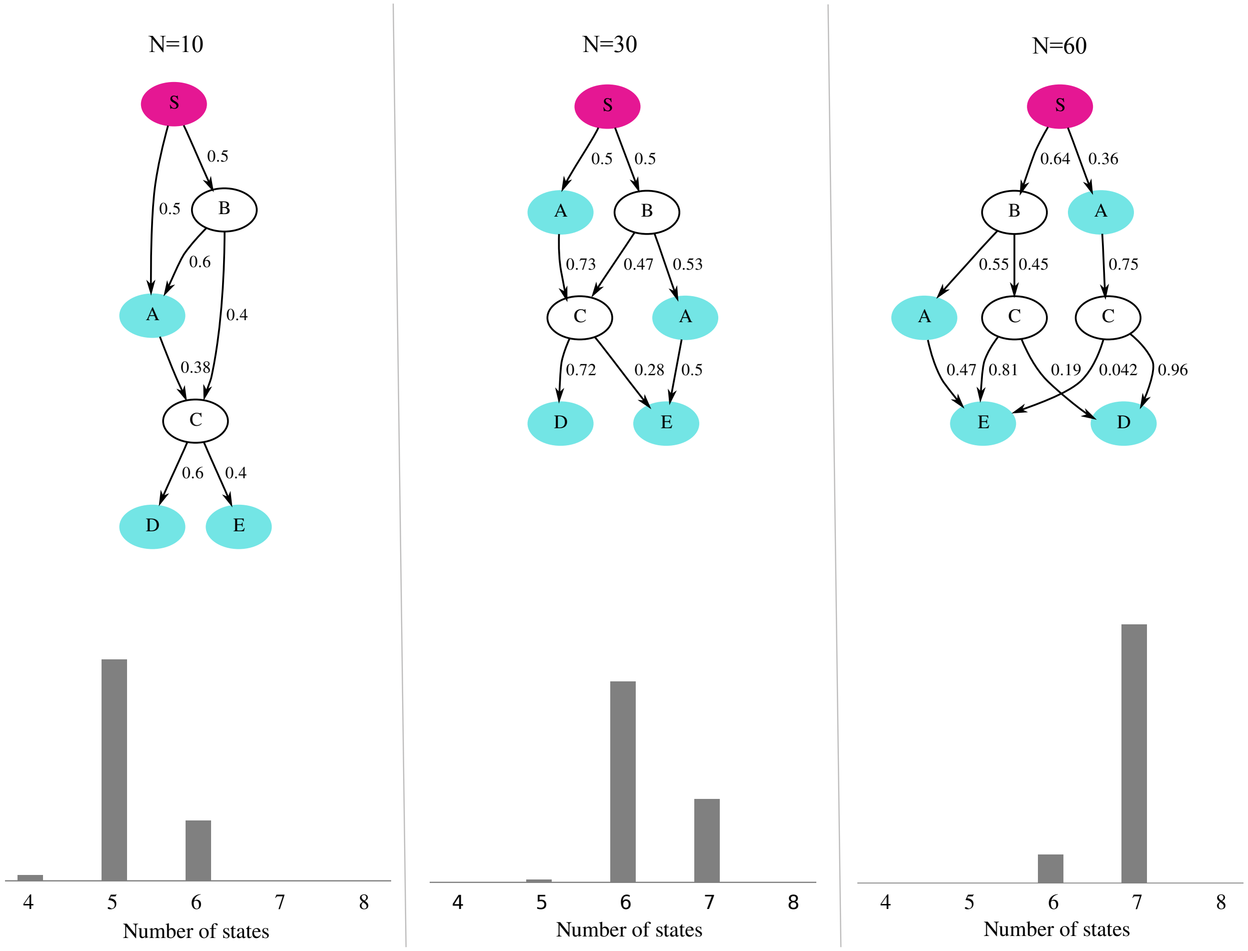


Fig. 4 Derived POMM for the example.

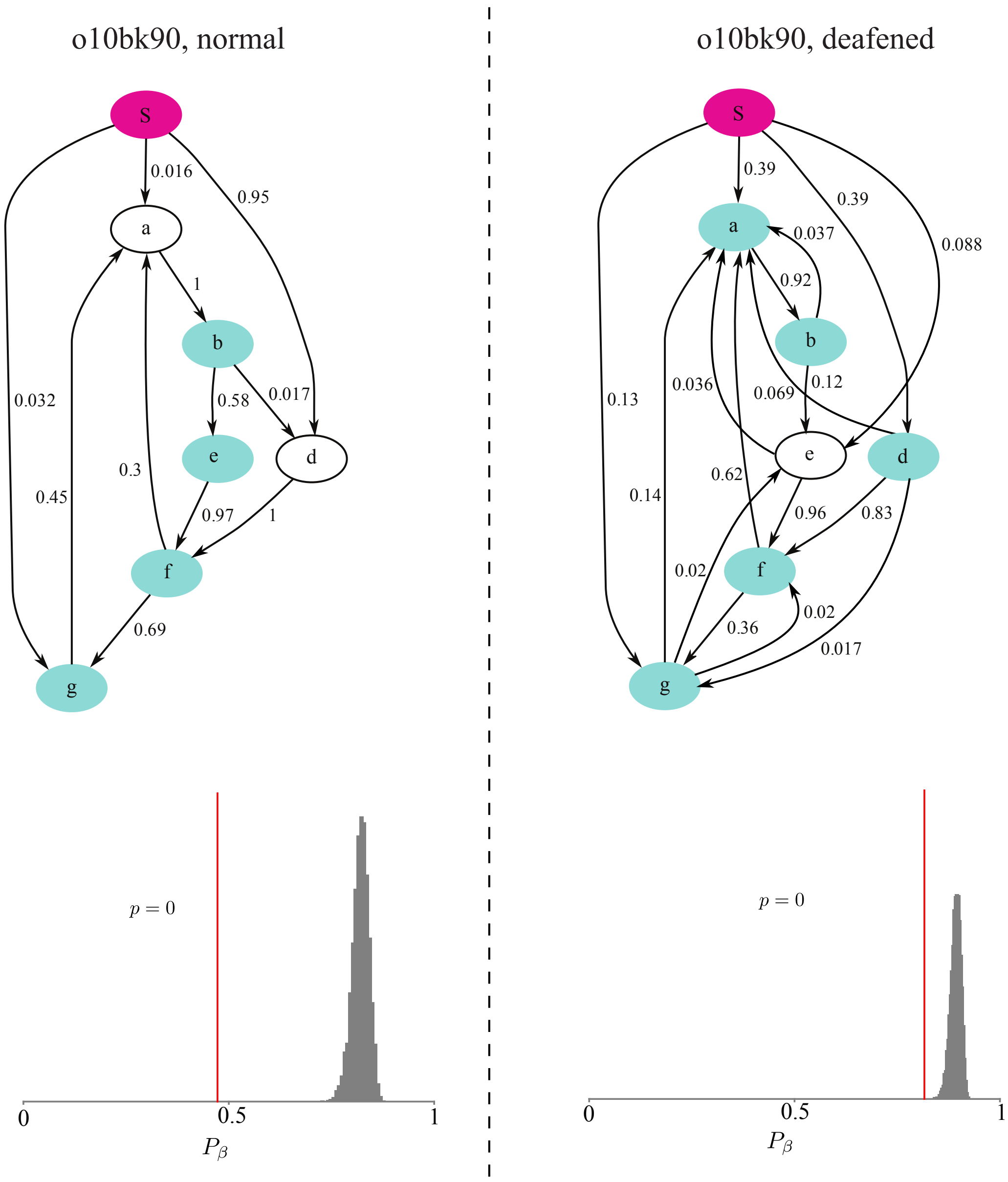


Fig. 5 Test of Markov model for bird o10bk90.

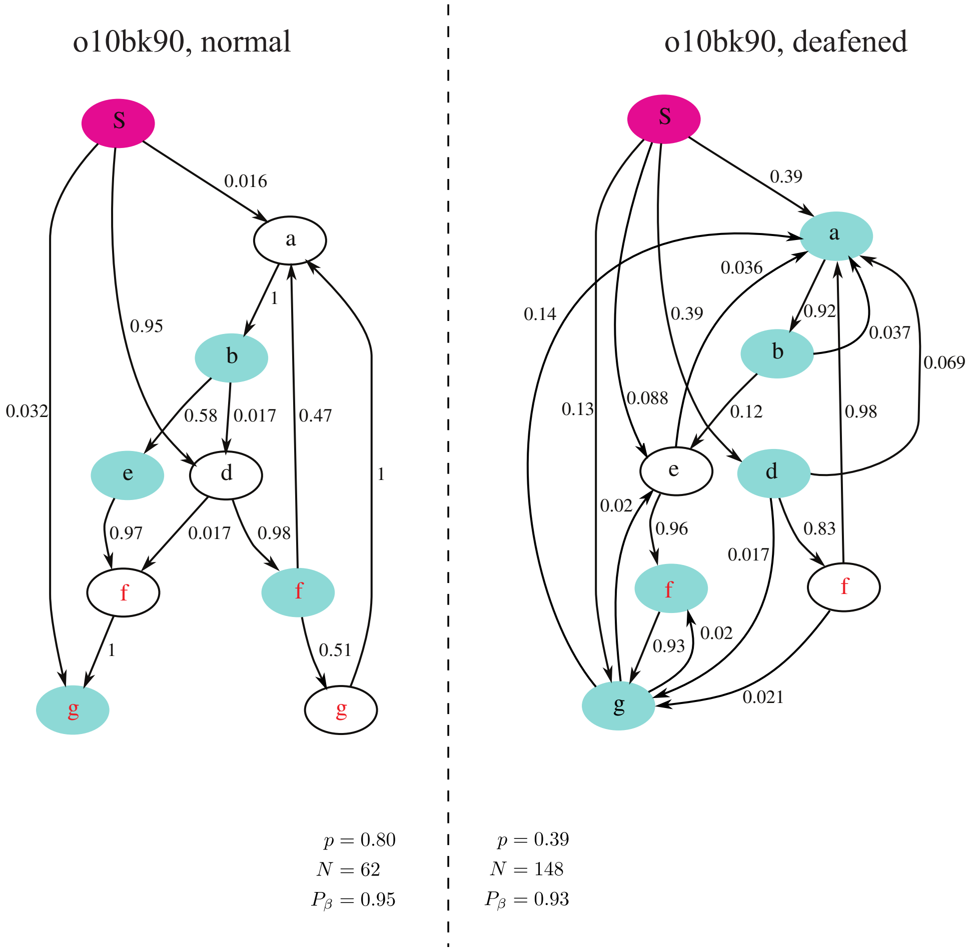


Fig. 6 POMM for bird o10bk90.

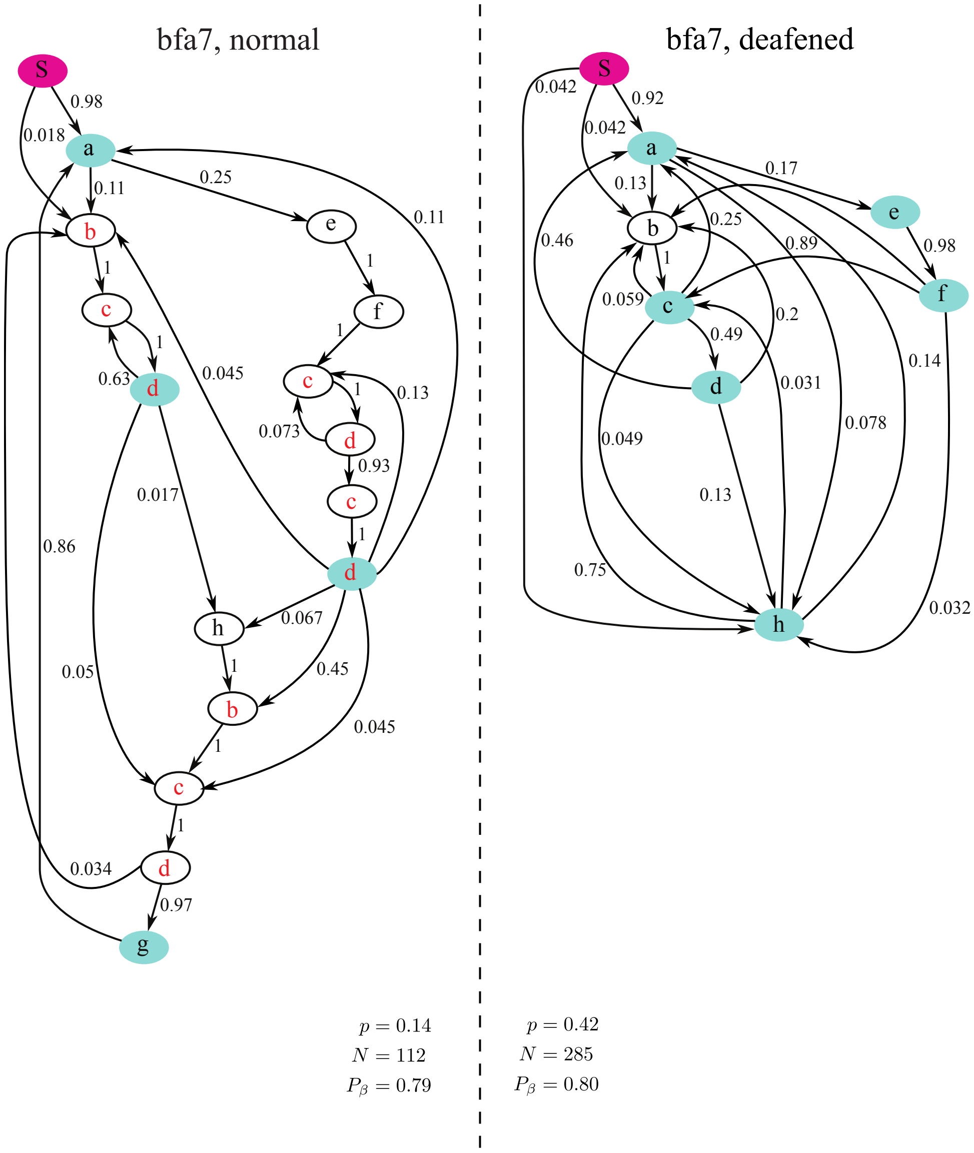


Fig. 7 POMM for bird bfa7 .

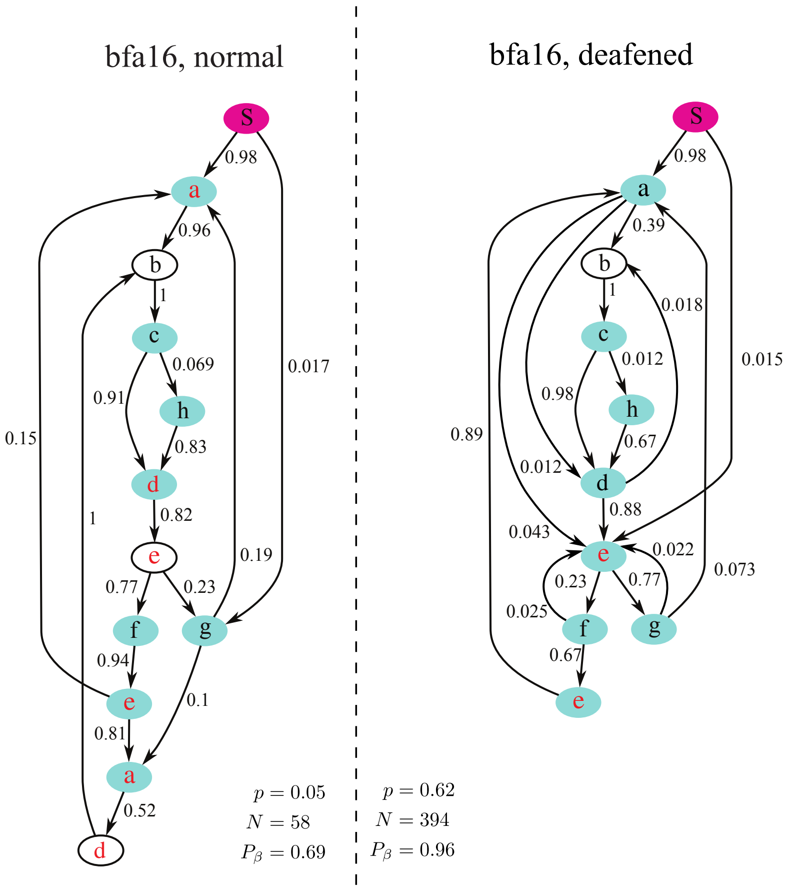


Fig. 8 POMM for bird bfa16.

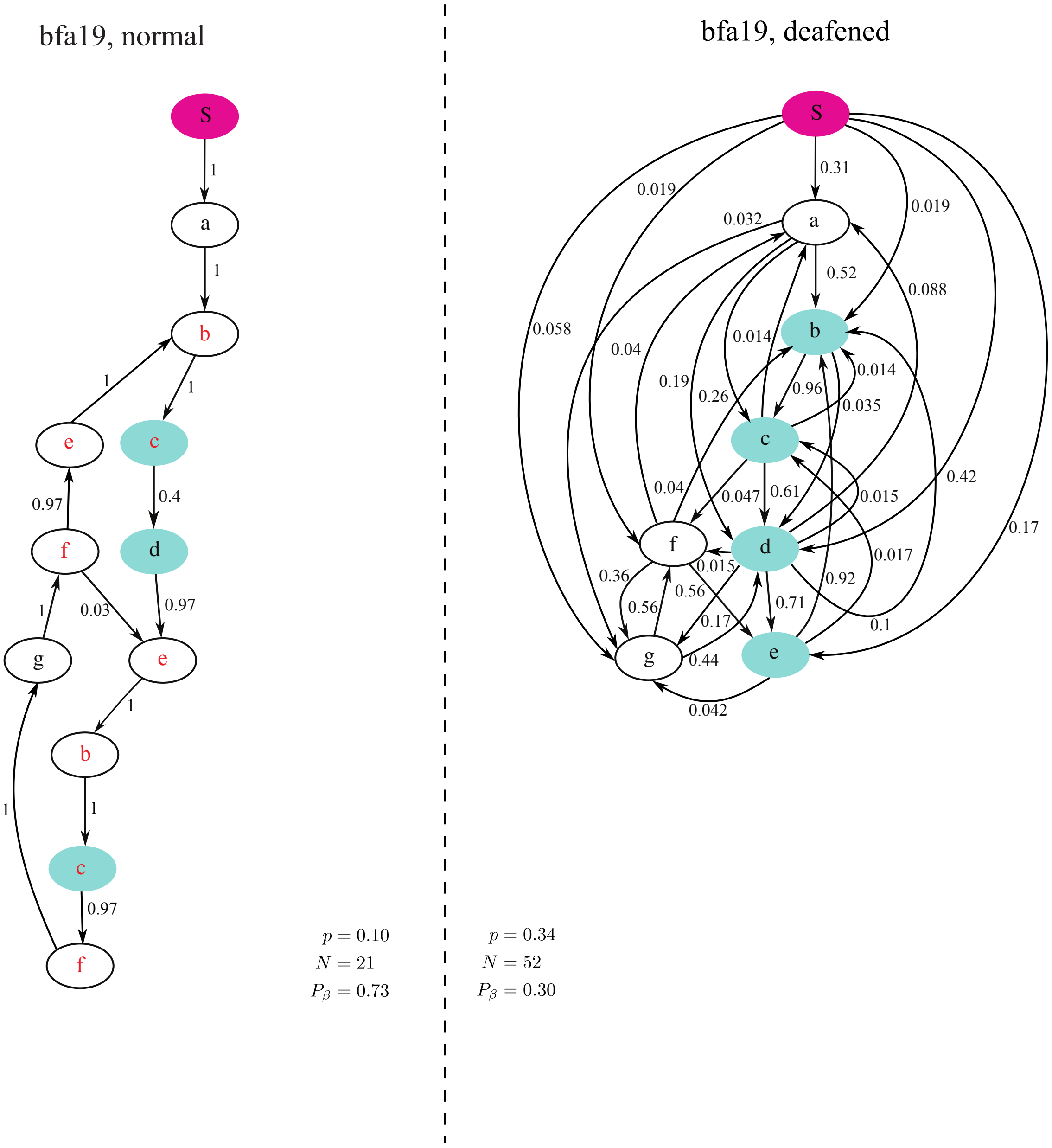


Fig. 9 POMM for bird bfa19.

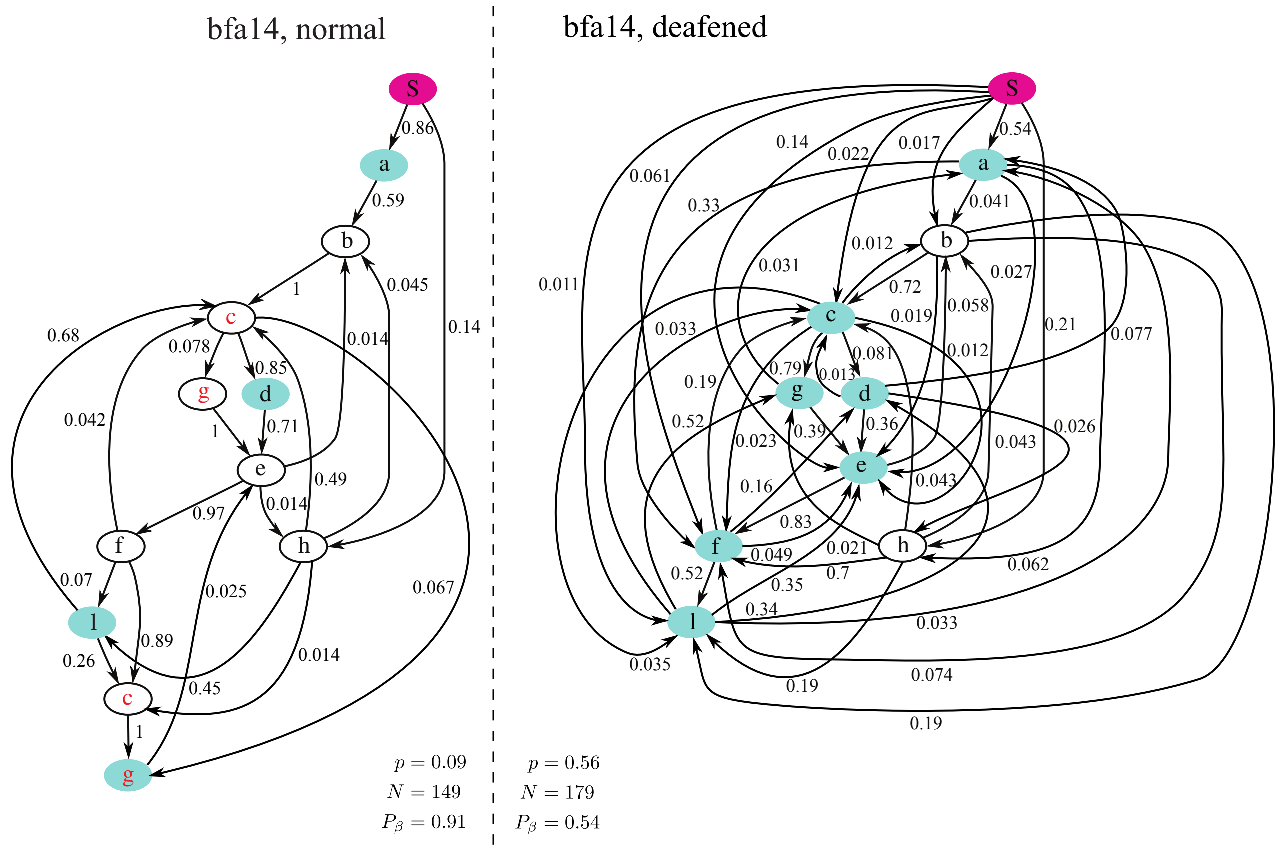
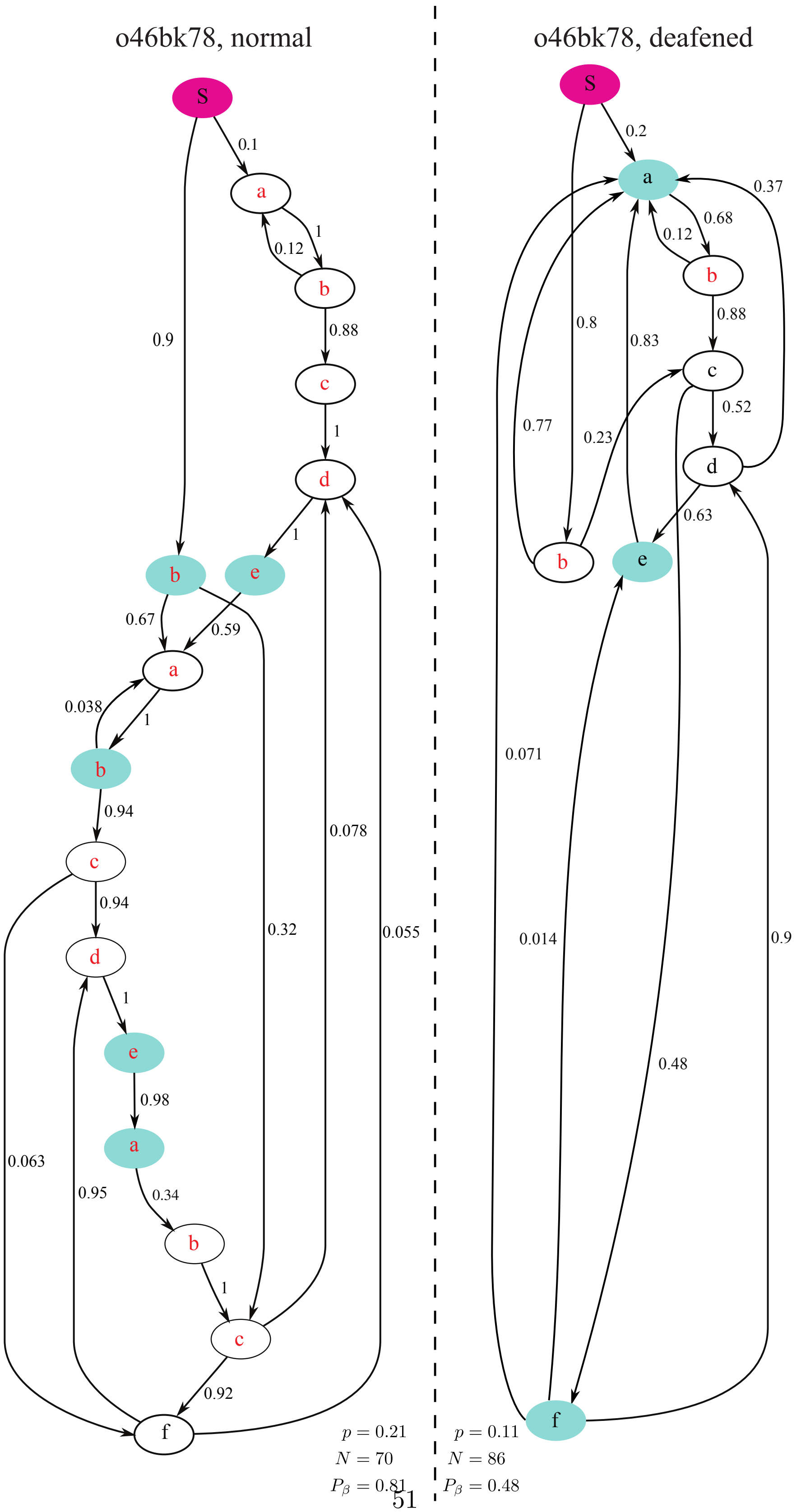


Fig. 10 POMM for bird bfa14.


o46bk78, normal

o46bk78, deafened

Fig. 11 POMM for bird o46bk78.

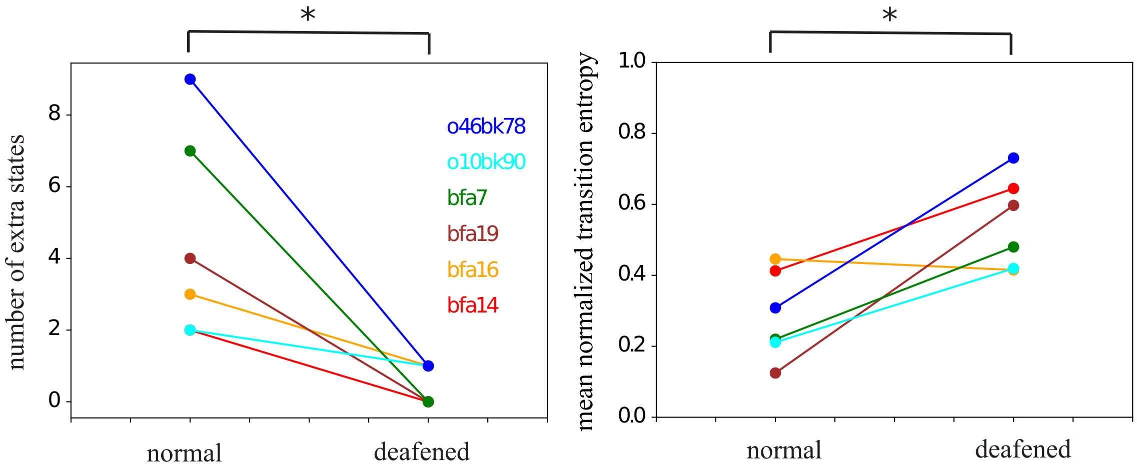
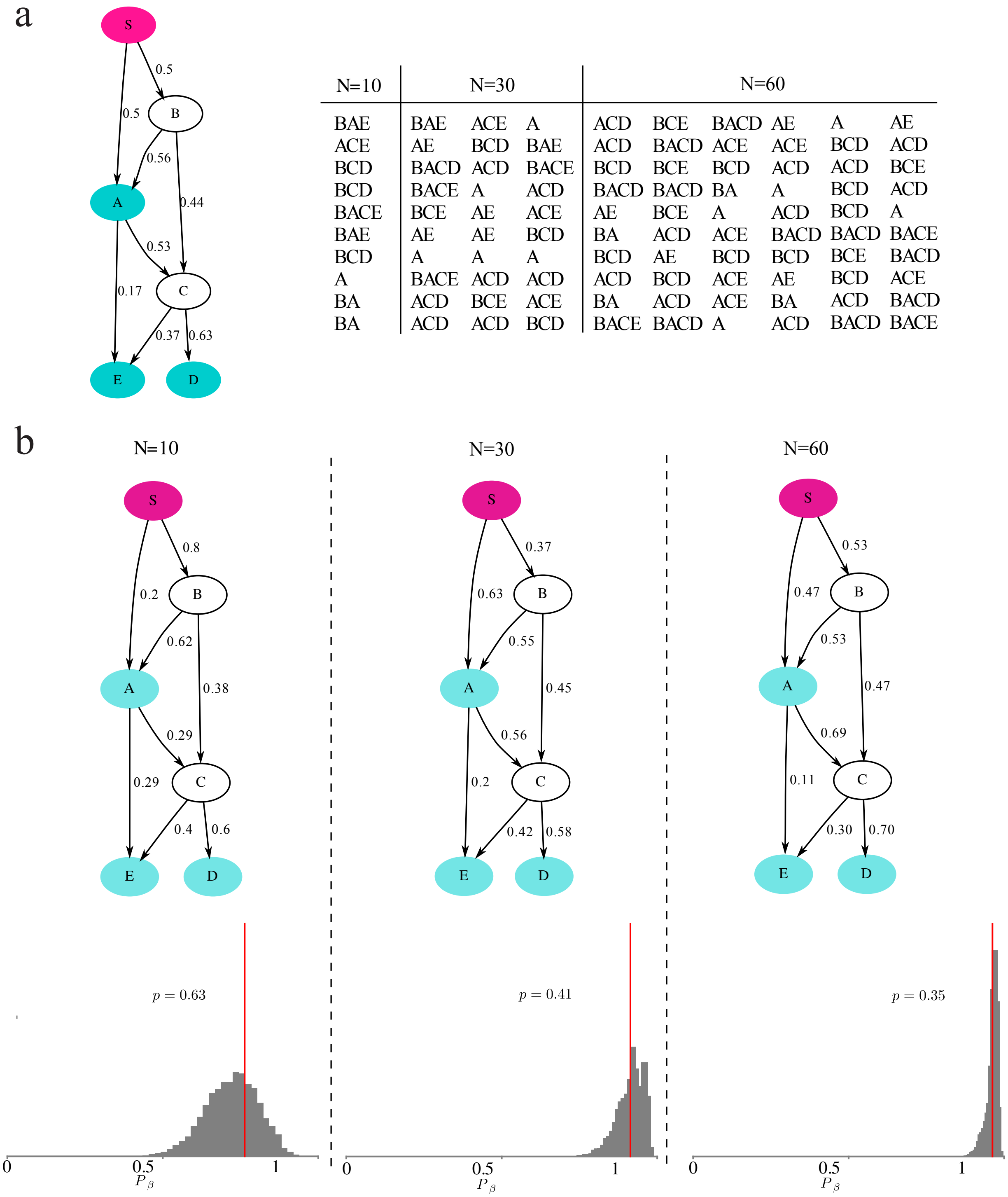


Fig. 12 Summary of the effects of deafening on POMM.



N=10	N=30			N=60					
BAE	BAE	ACE	A	ACD	BCE	BACD	AE	A	AE
ACE	AE	BCD	BAE	ACD	BACD	ACE	ACE	BCD	ACD
BCD	BACD	ACD	BACE	BCD	BCE	BCD	ACD	ACD	BCE
BCD	BACE	A	ACD	BACD	BACD	BA	A	BCD	ACD
BACE	BCE	AE	ACE	AE	BCE	A	ACD	BCD	A
BAE	AE	AE	BCD	BA	ACD	ACE	BACD	BACD	BACE
BCD	A	A	A	BCD	AE	BCD	BCD	BCE	BACD
A	BACE	ACD	ACD	ACD	BCD	ACE	AE	BCD	ACE
BA	ACD	BCE	ACE	BA	ACD	ACE	BA	ACD	BACD
BA	ACD	ACD	BCD	BACE	BACD	A	ACD	BACD	BACE

Fig. S1 (Supplementary) Statistical test of Markov model.

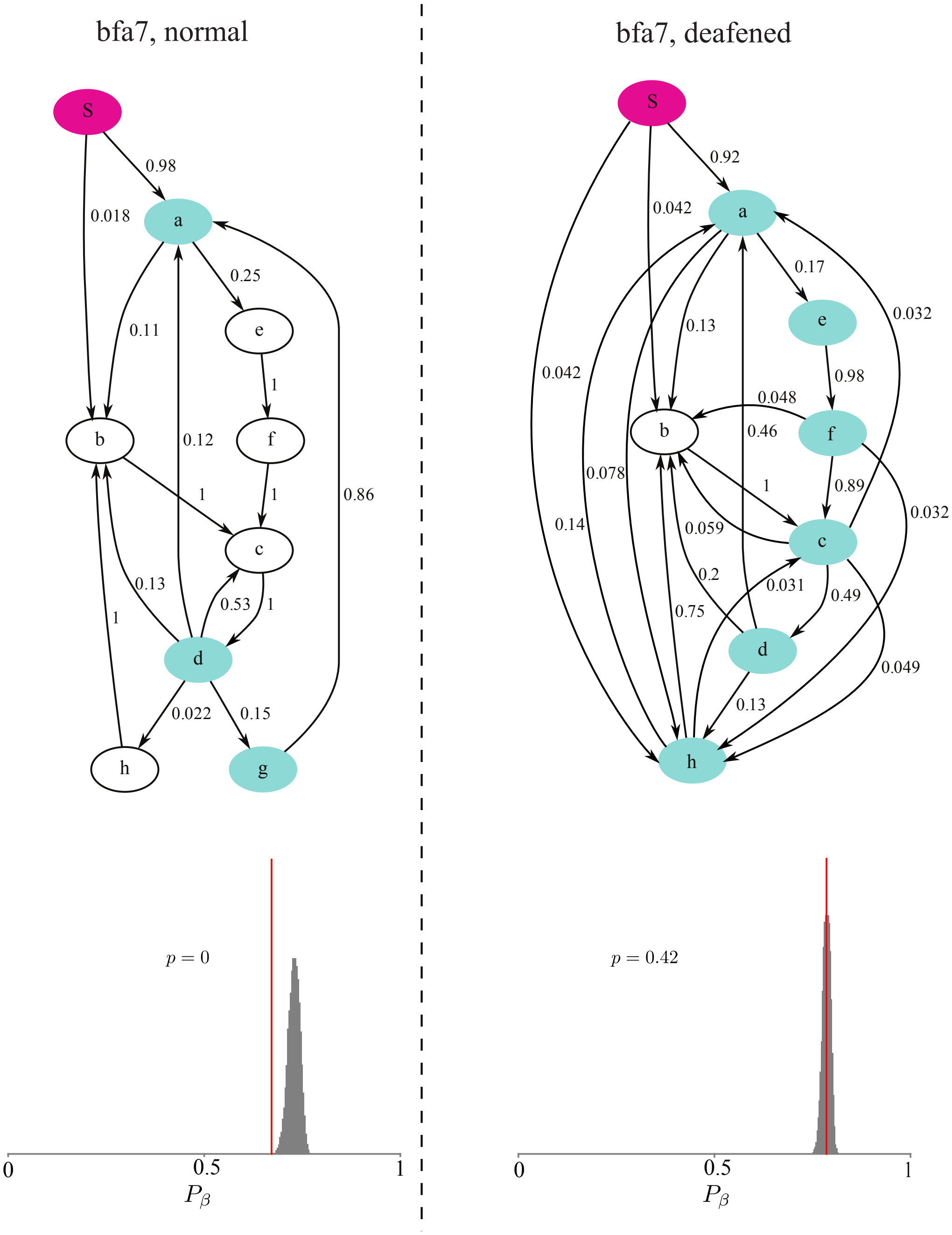


Fig. S2 (Supplementary) Test of Markov model for bird bfa7 .

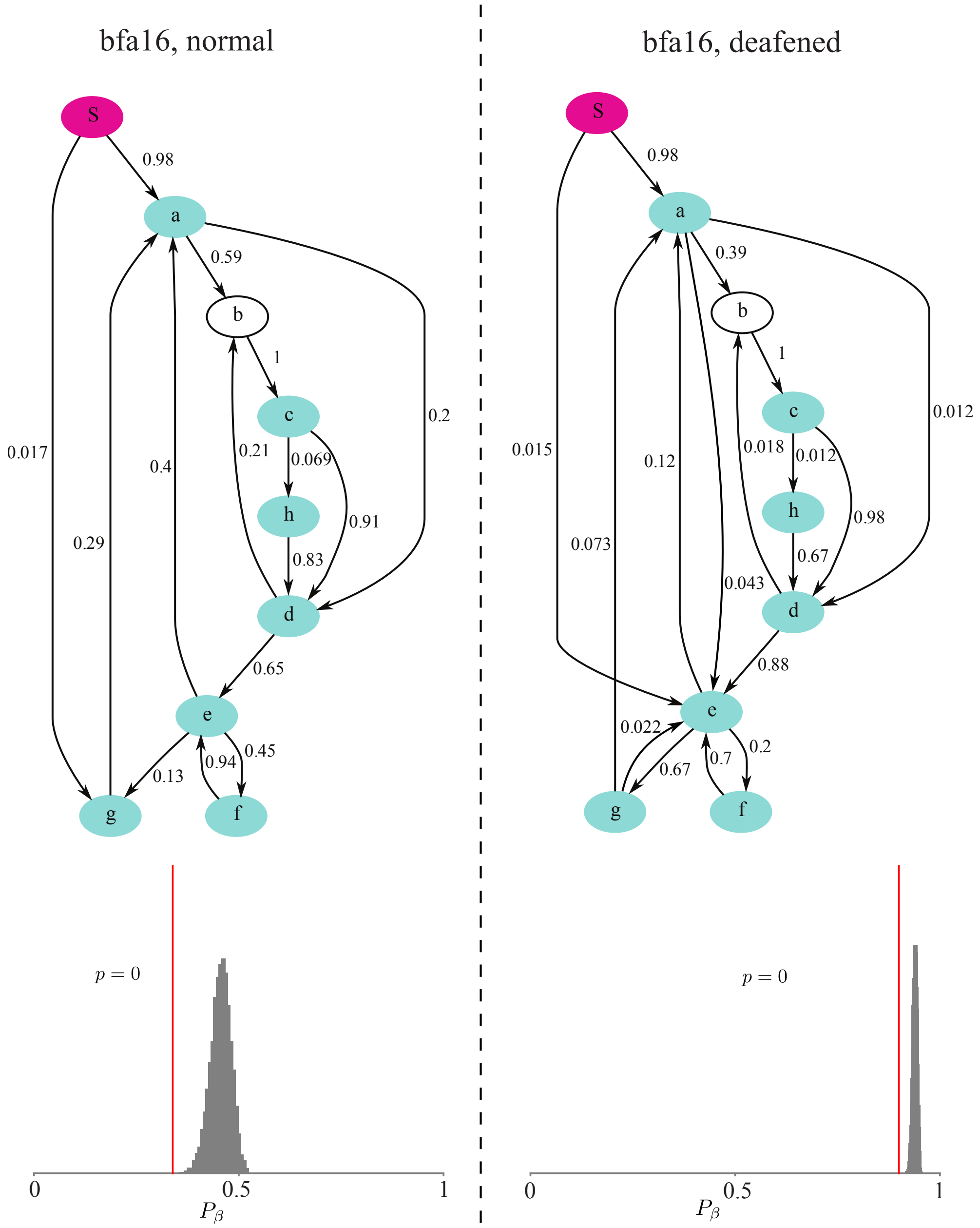


Fig. S3 (Supplementary) Test of Markov model for bird bfa16.

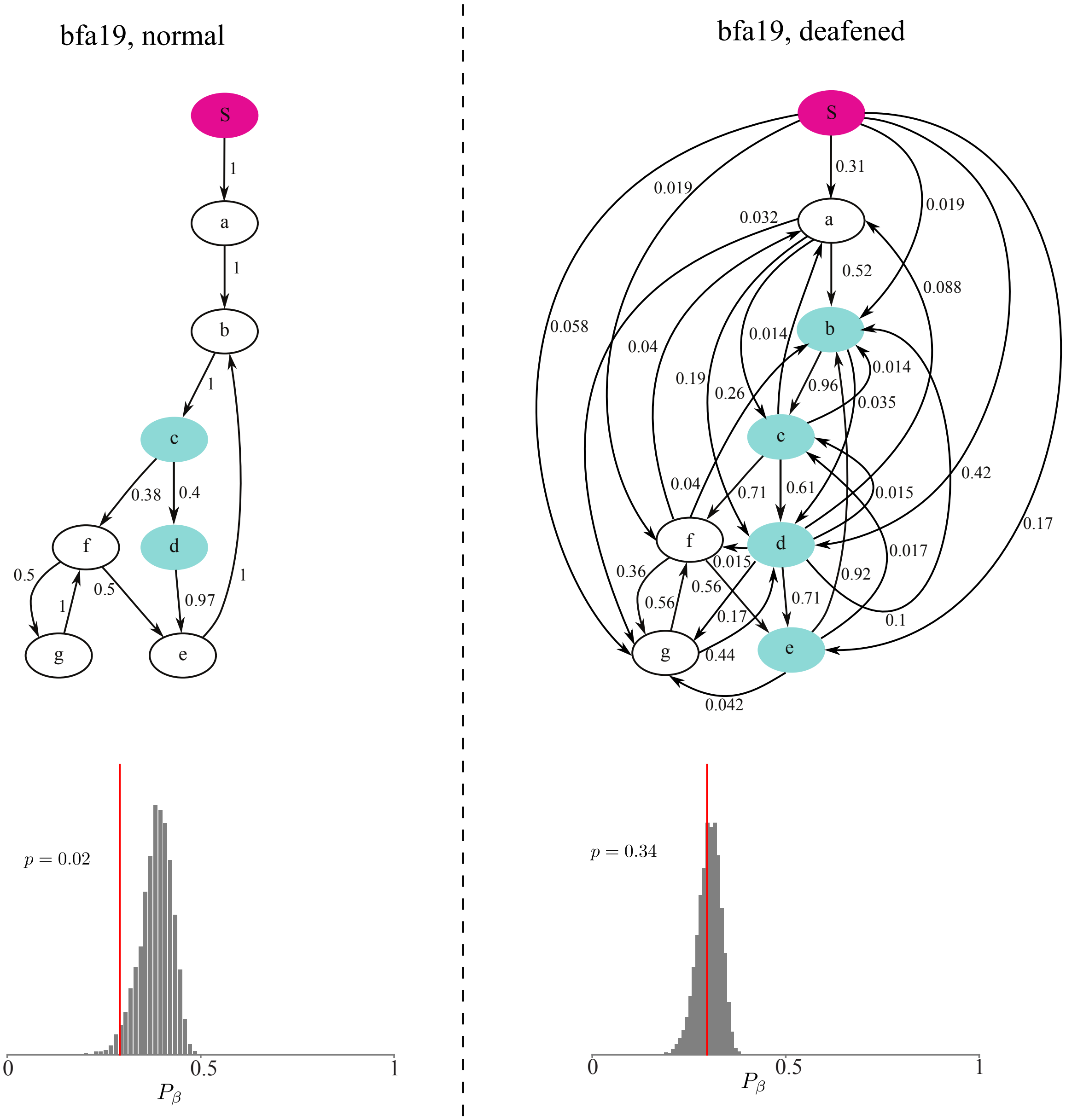


Fig. S4 (Supplementary) Test of Markov model for bird bfa19.

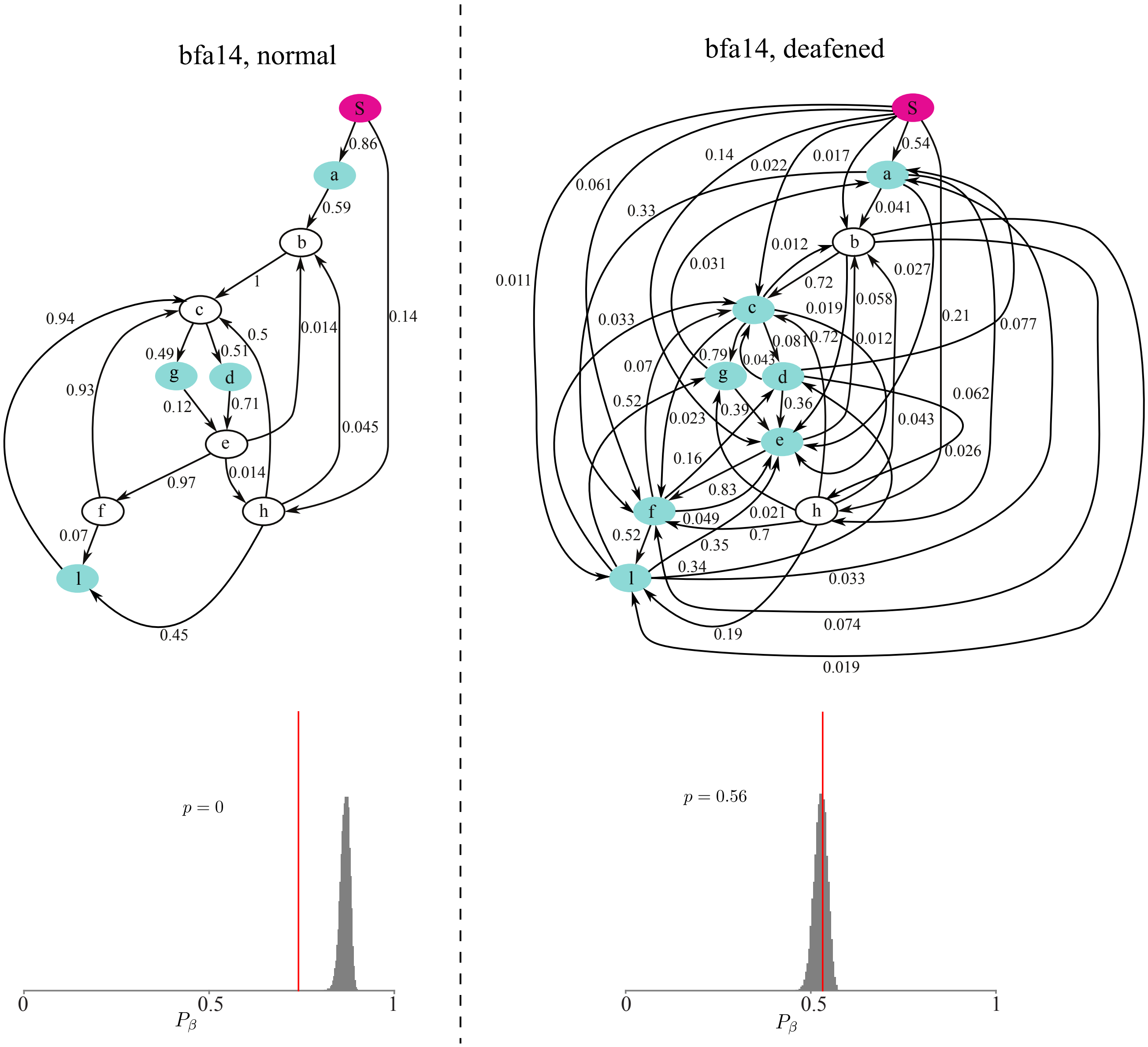
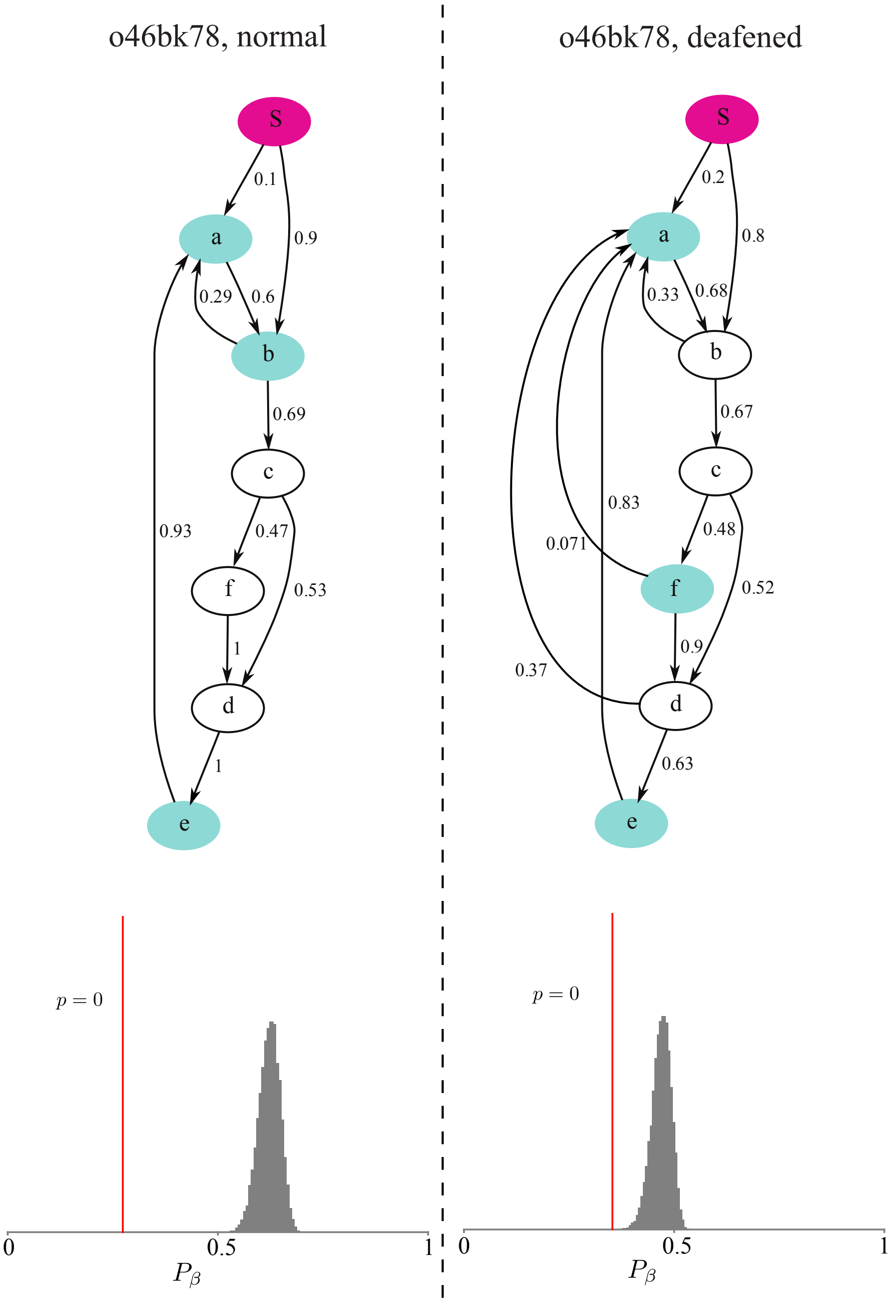


Fig. S5 (Supplementary) Test of Markov model for bird bfa14.



794

796 **Fig. S6 (Supplementary) Test of Markov model for bird o46bk78.**

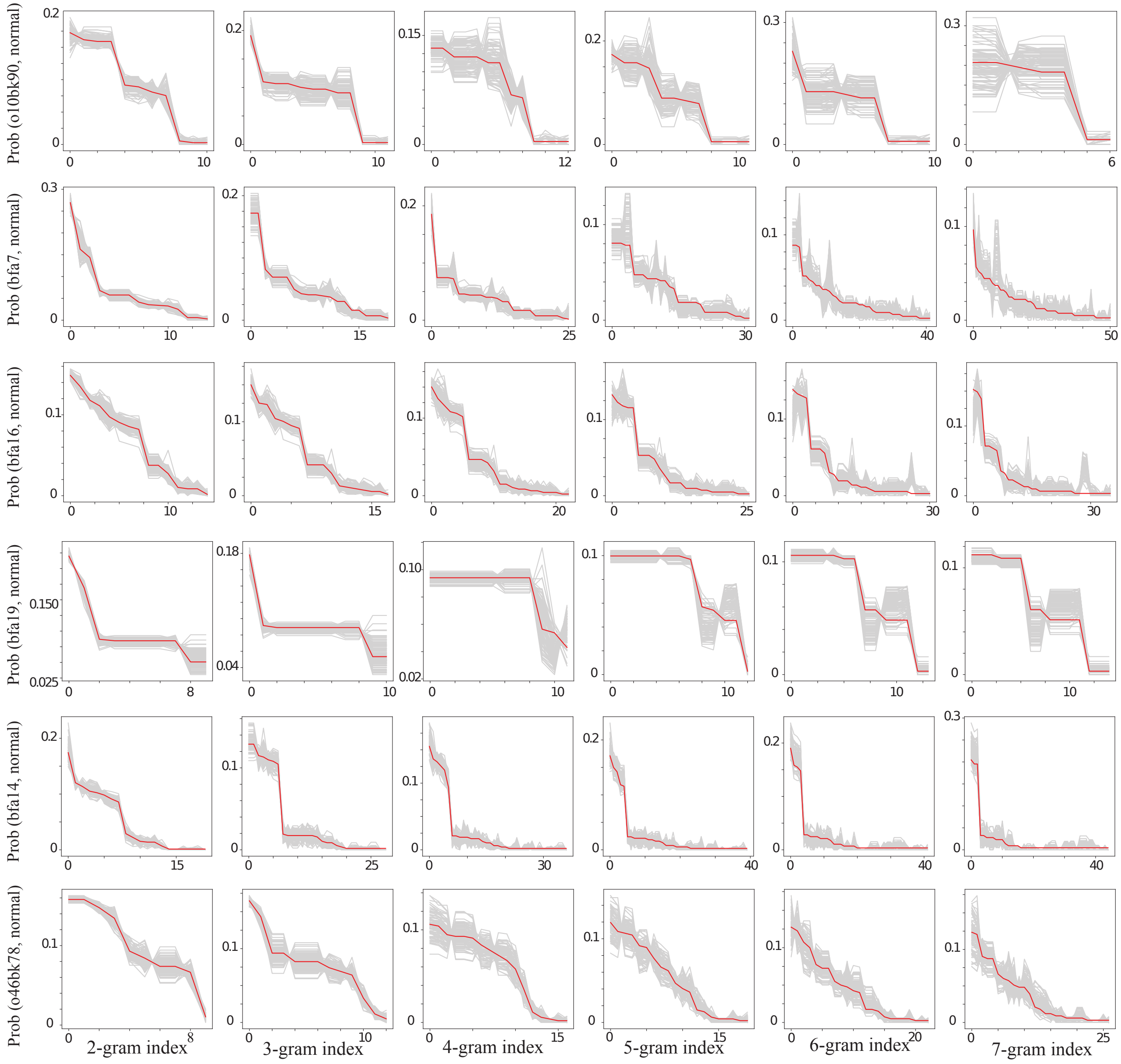


Fig. S7 (Supplementary) Comparisons of N-gram distributions in normal hearing condition.

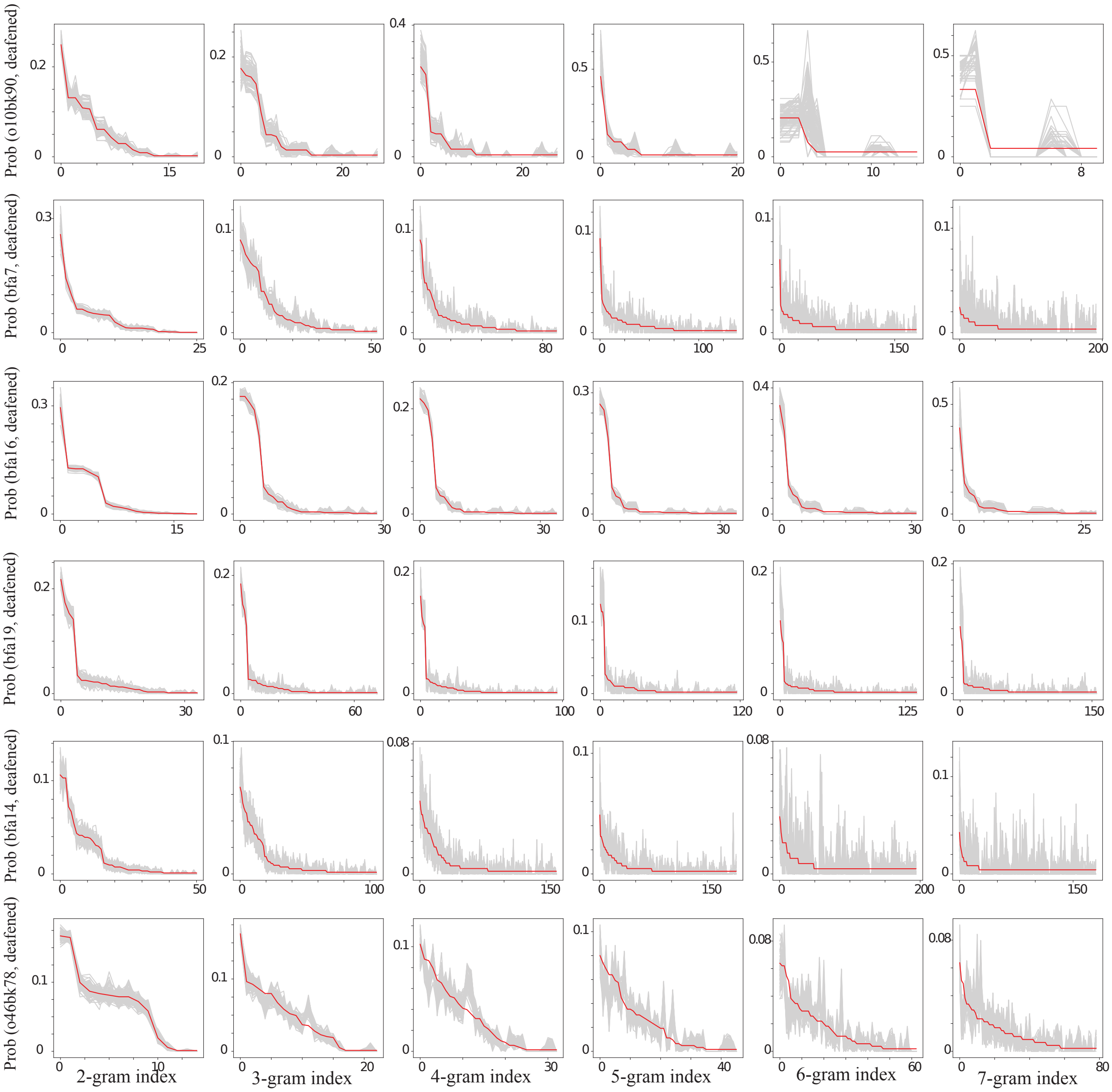


Fig. S8 (Supplementary) Comparisons of N-gram distributions after deafeninig.